Silicon photonics and memories

Vladimir Stojanović

Integrated Systems Group, RLE/MTL

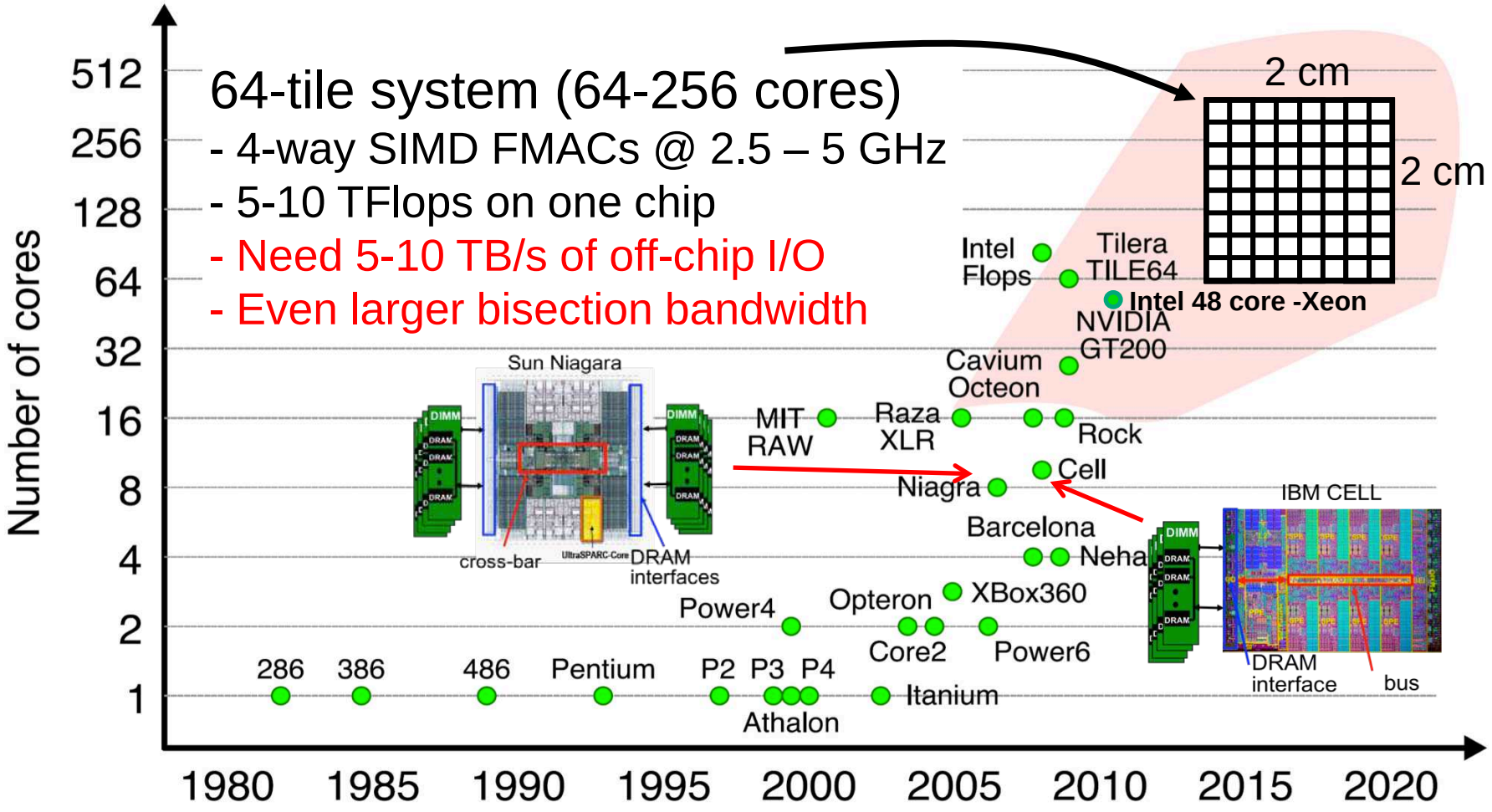
MIT

Acknowledgments

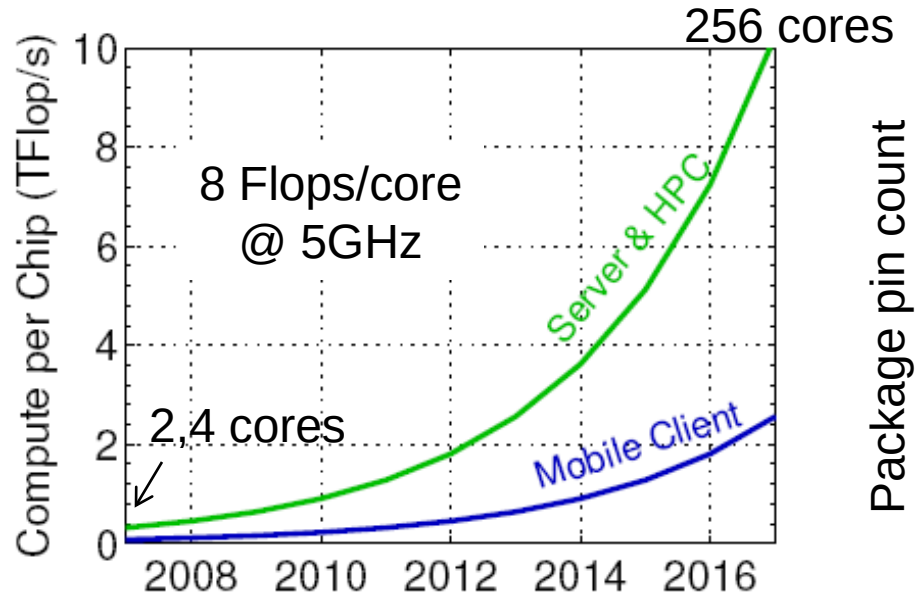
- Krste Asanović, Christopher Batten, Ajay Joshi
  - Scott Beamer, Chen Sun, Yon-Jin Kwon, Imran Shamim
- Rajeev Ram, Milos Popovic, Franz Kaertner, Judy Hoyt, Henry Smith, Erich Ippen
  - Hanqin Li, Charles Holzwarth
  - Jason Orcutt, Anatoly Khilo, Jie Sun, Cheryl Sorace, Eugen Zgraggen
  - Michael Georgas, Jonathan Leu, Ben Moss

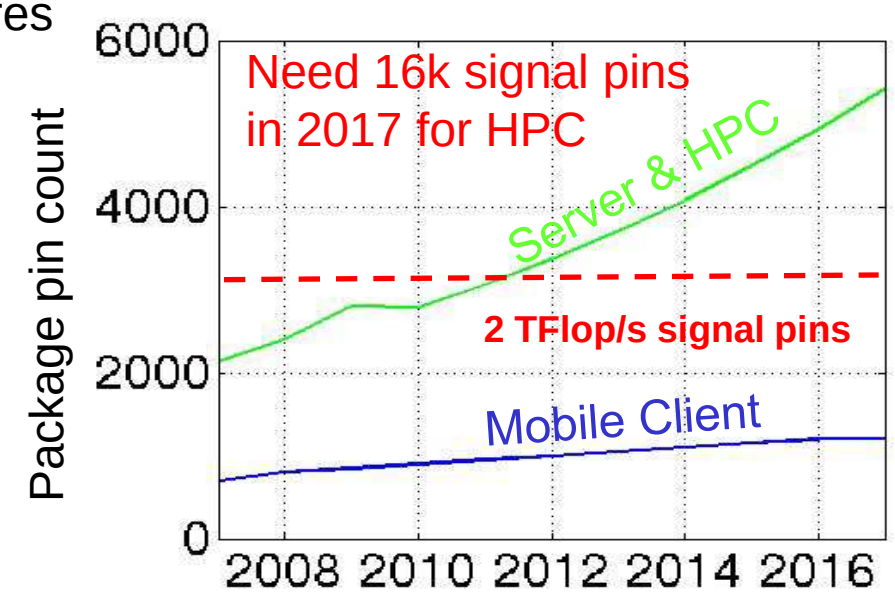
- Dr. Jag Shah – DARPA MTO
- Texas Instruments
- Intel Corporation

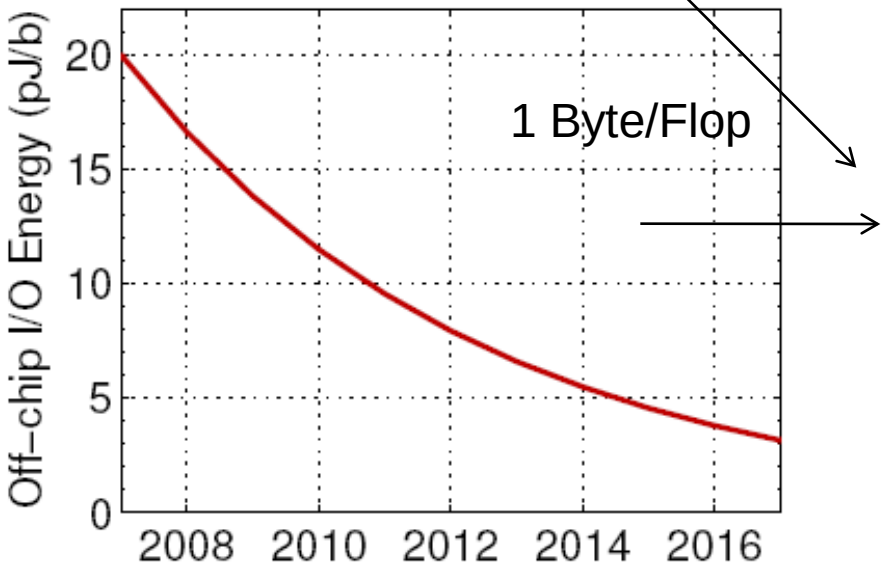
Processors scaling to manycore systems

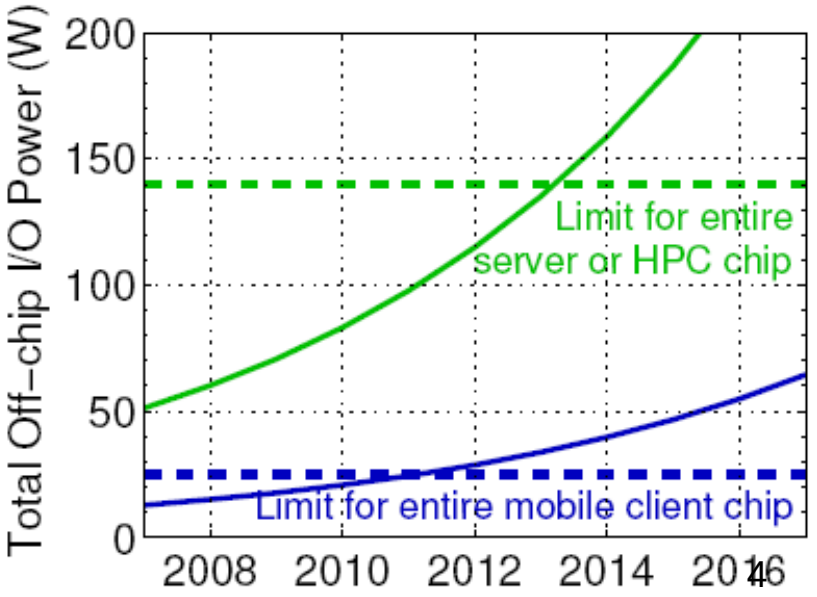


Bandwidth, pin count and power scaling









Electrical Baseline in 2016

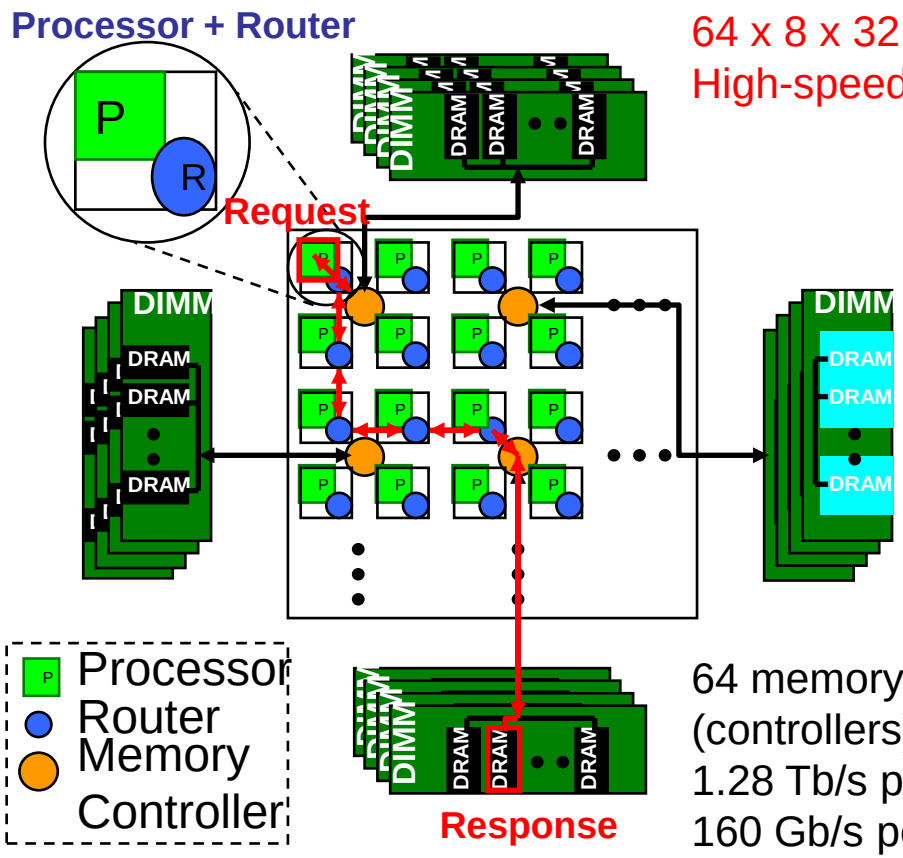
Node Board
10 TFlop/s
512 GB DRAM
80 Tb/s mem BW

CPU Power 1kW -> 100W
Energy-efficiency
100 pJ/Flop -> 10pJ/Flop

Cross-chip
I/O
Compute

Memory Power 1kW

Cross-chip
I/O
Activate

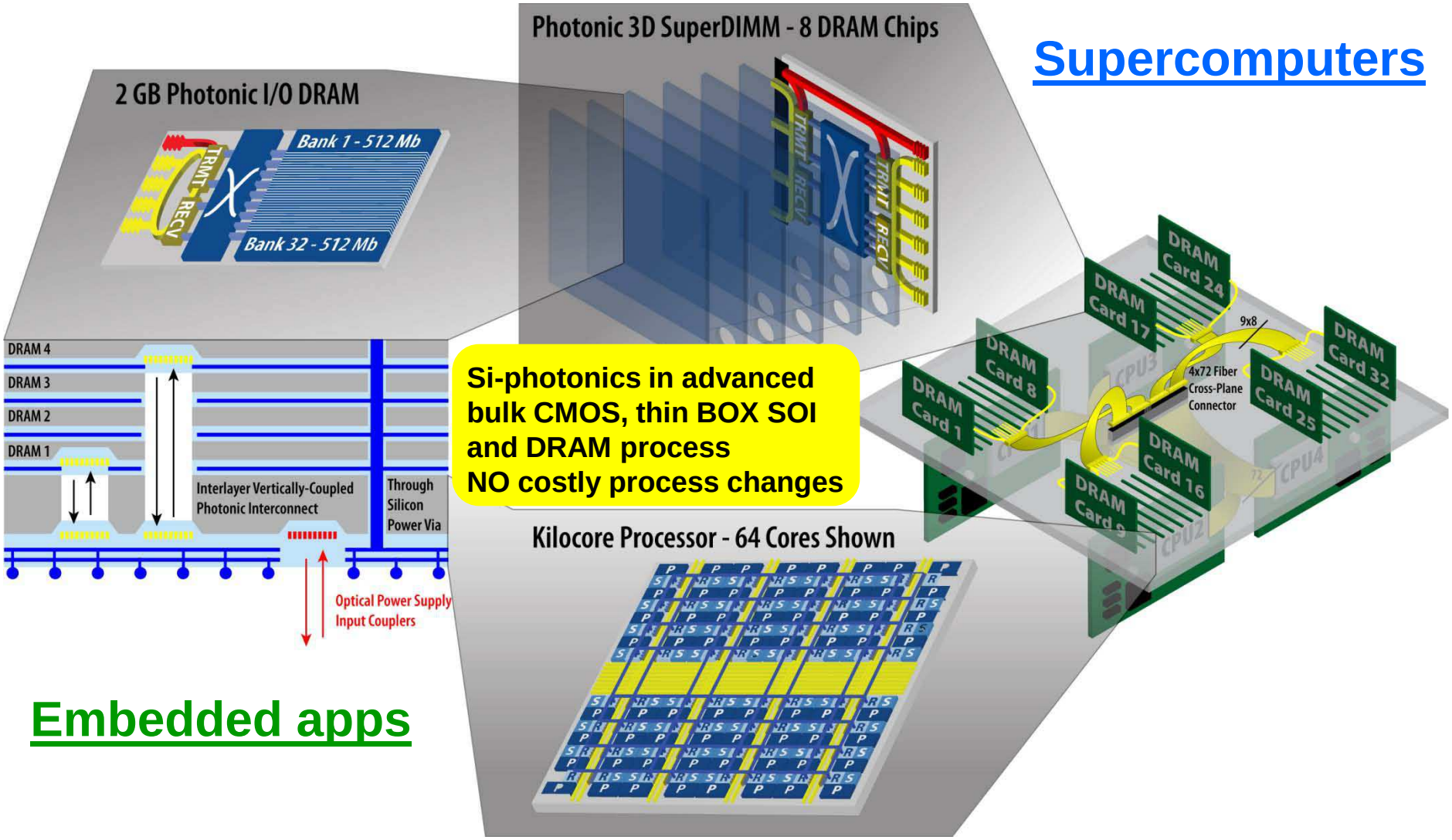


CPU
64 x 8 x 32 = 16k
High-speed signal pins

64 memory channels (controllers)
1.28 Tb/s per controller
160 Gb/s per chip
(16 x 10 Gb/s) @ 5pJ/b

512 x 1GB DRAM chips
8 chips per DIMM
1DIMM per memory channel
Need at least 16 banks/chip to sustain BW

Monolithic CMOS-Photonics in Computer Systems



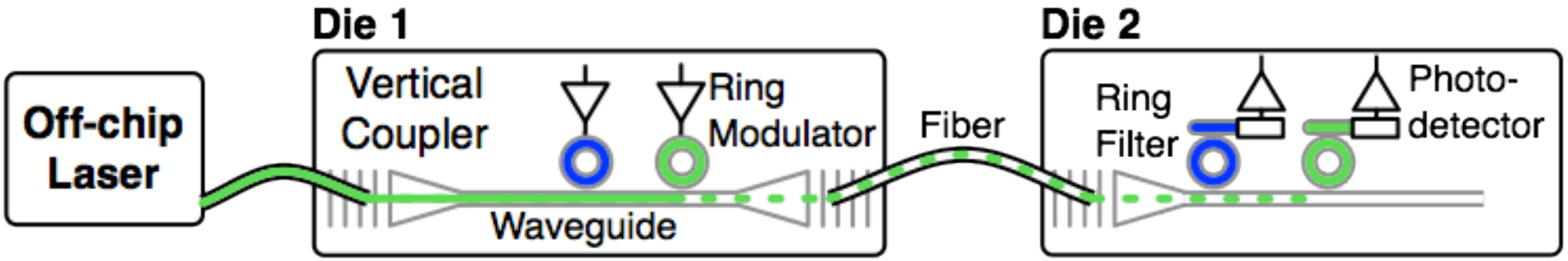
Supercomputers

Embedded apps

Bandwidth density – need dense WDM

Energy-efficiency – need monolithic integration

CMOS photonics density and energy advantage

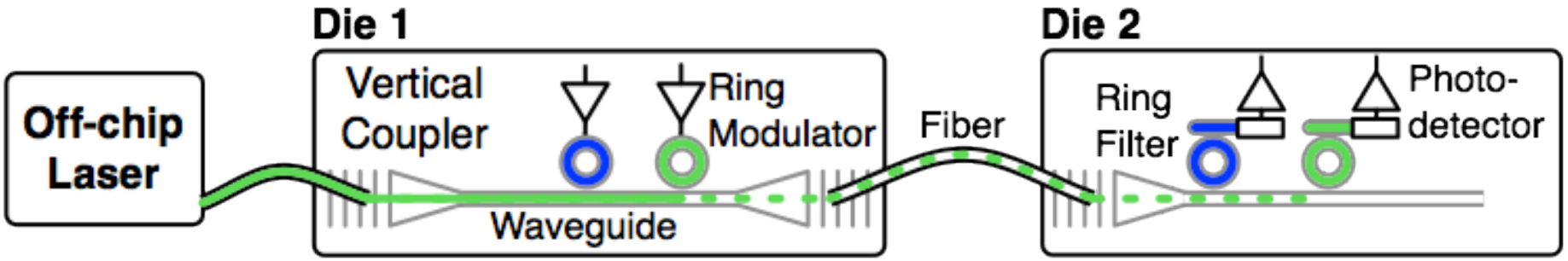


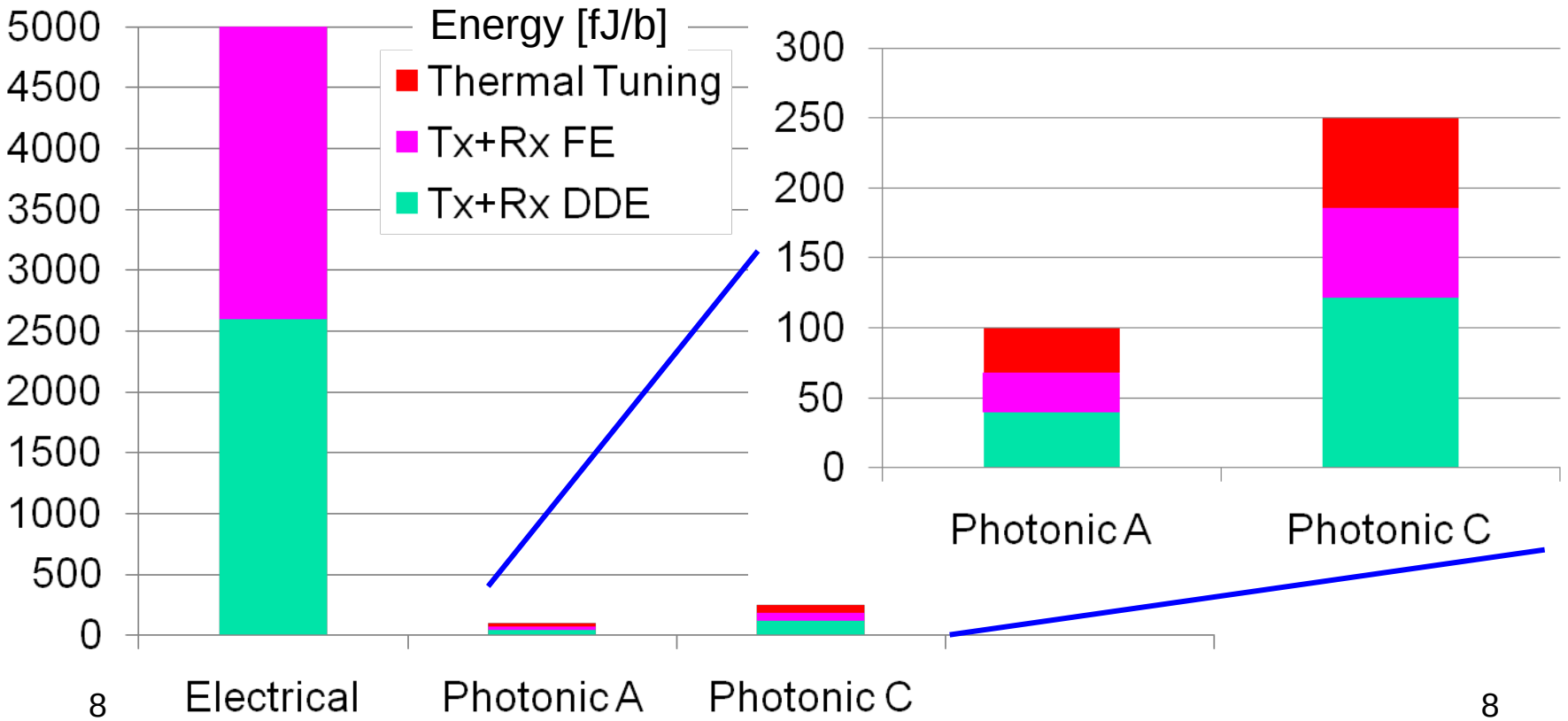
Metric	Energy (pJ/b)	Bandwidth density (Gb/s/μ)
Global on-chip photonic link	0.1-0.25	160-320
Global on-chip optimally repeated electrical link	1	5
Off-chip photonic link (100 μ coupler pitch)	0.1-0.25	6-13
Off-chip electrical SERDES (100 μ pitch)	5	0.1

Assuming 128 10Gb/s wavelengths on each waveguide

But, need to keep links fully utilized …

Fixed and static energy increase at low link utilization !

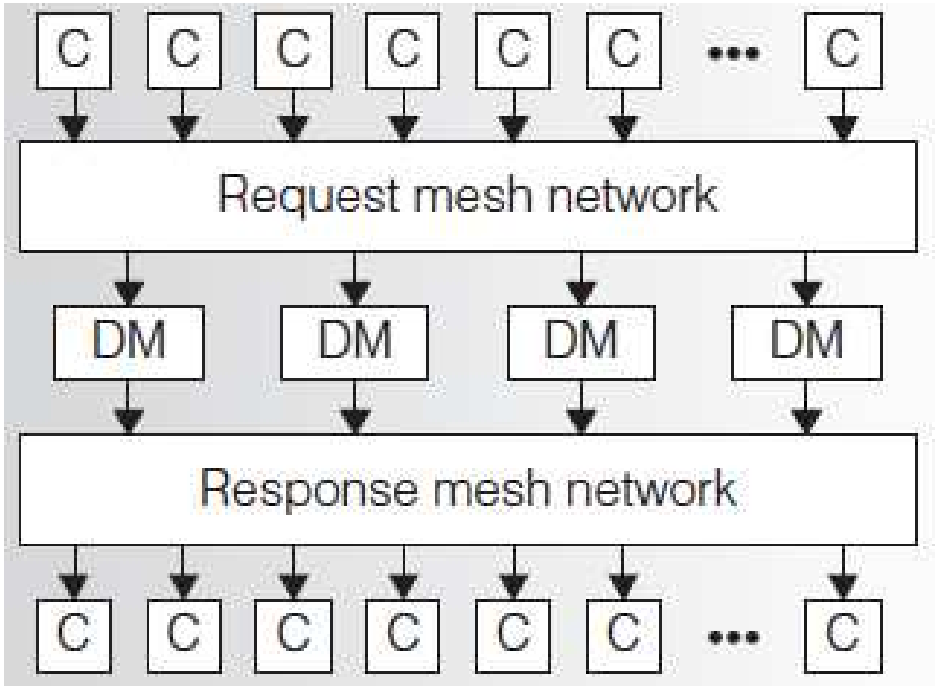


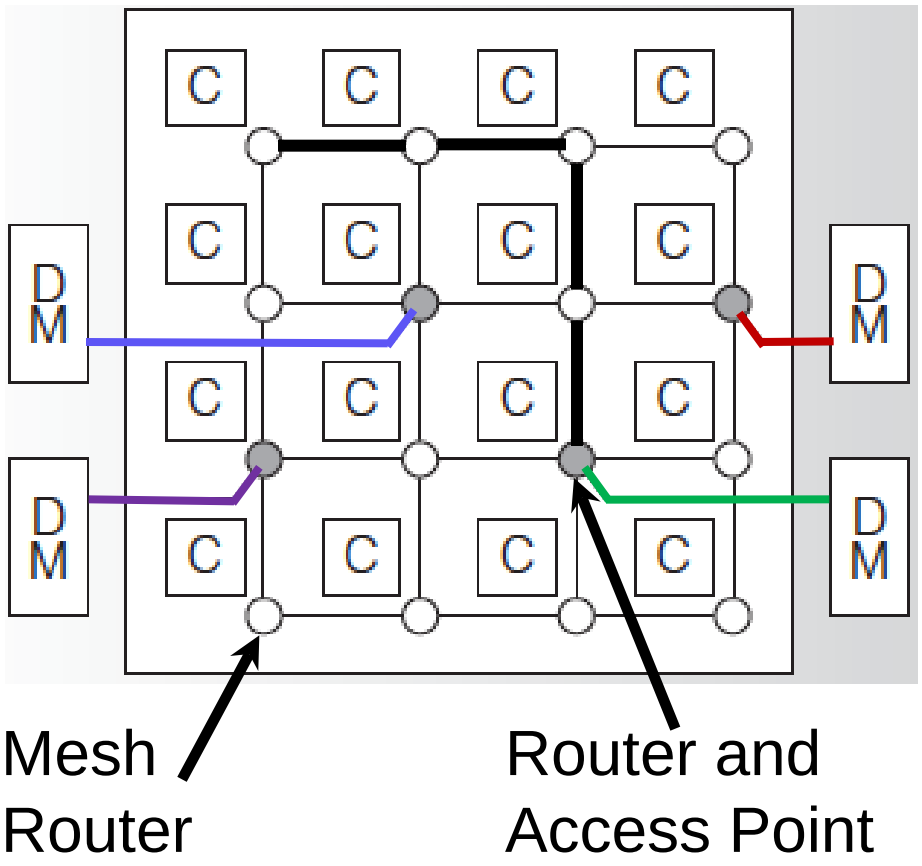


8

Core-to-Memory network: Electrical baseline

C = Core, DM = DRAM Module

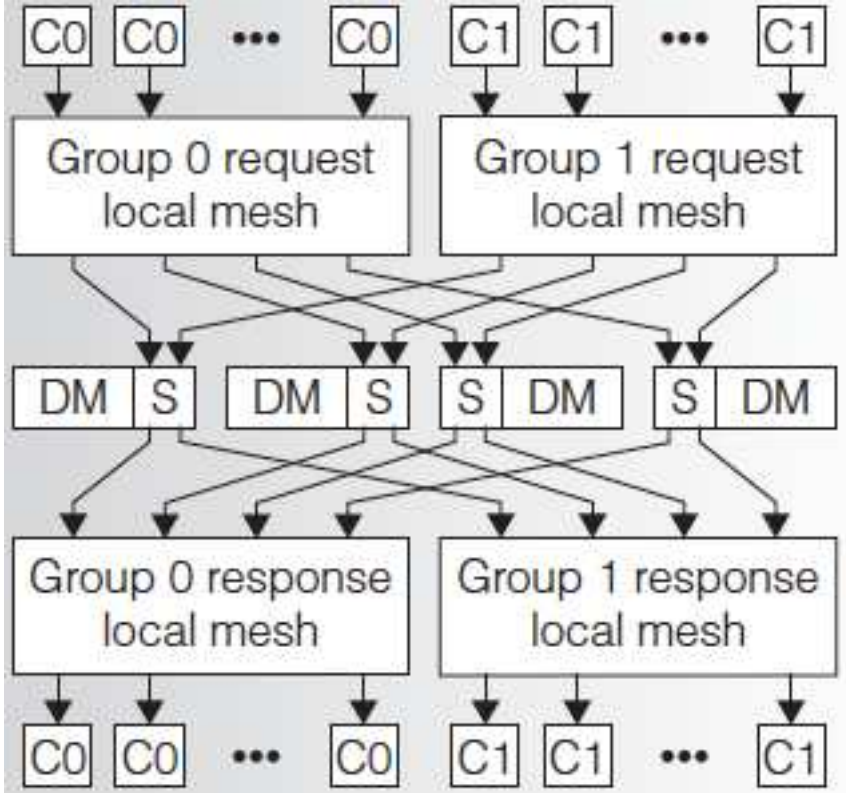


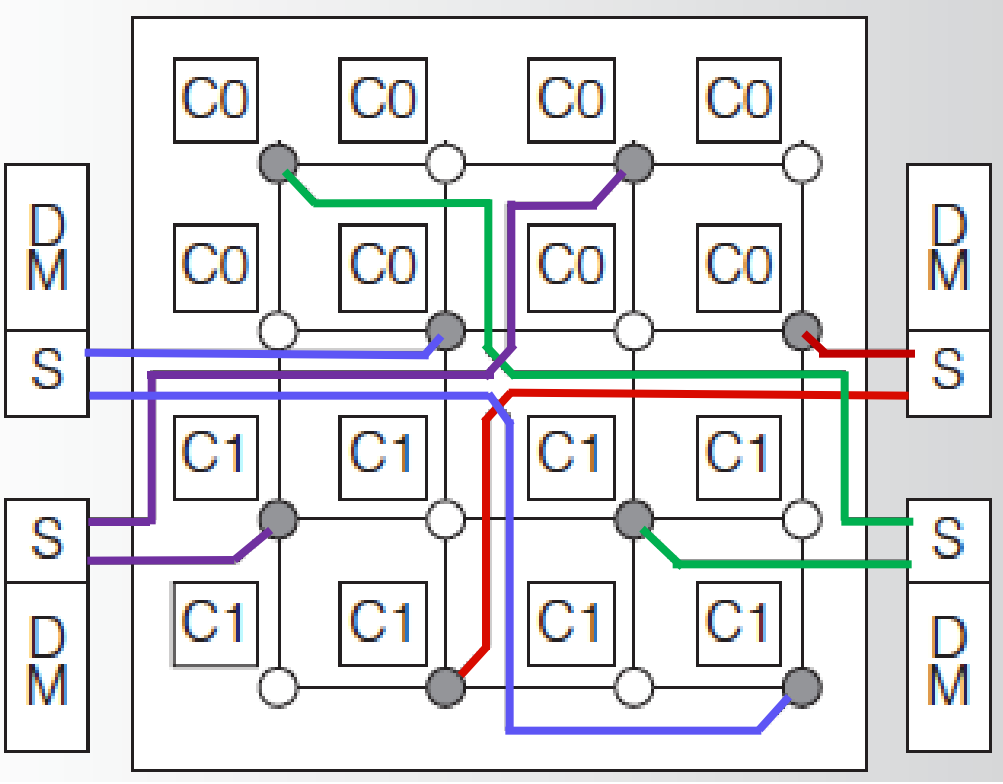


- Both cross-chip and I/O costly

Aggregation with Optical LMGS* network

* Local Meshes to Global Switches





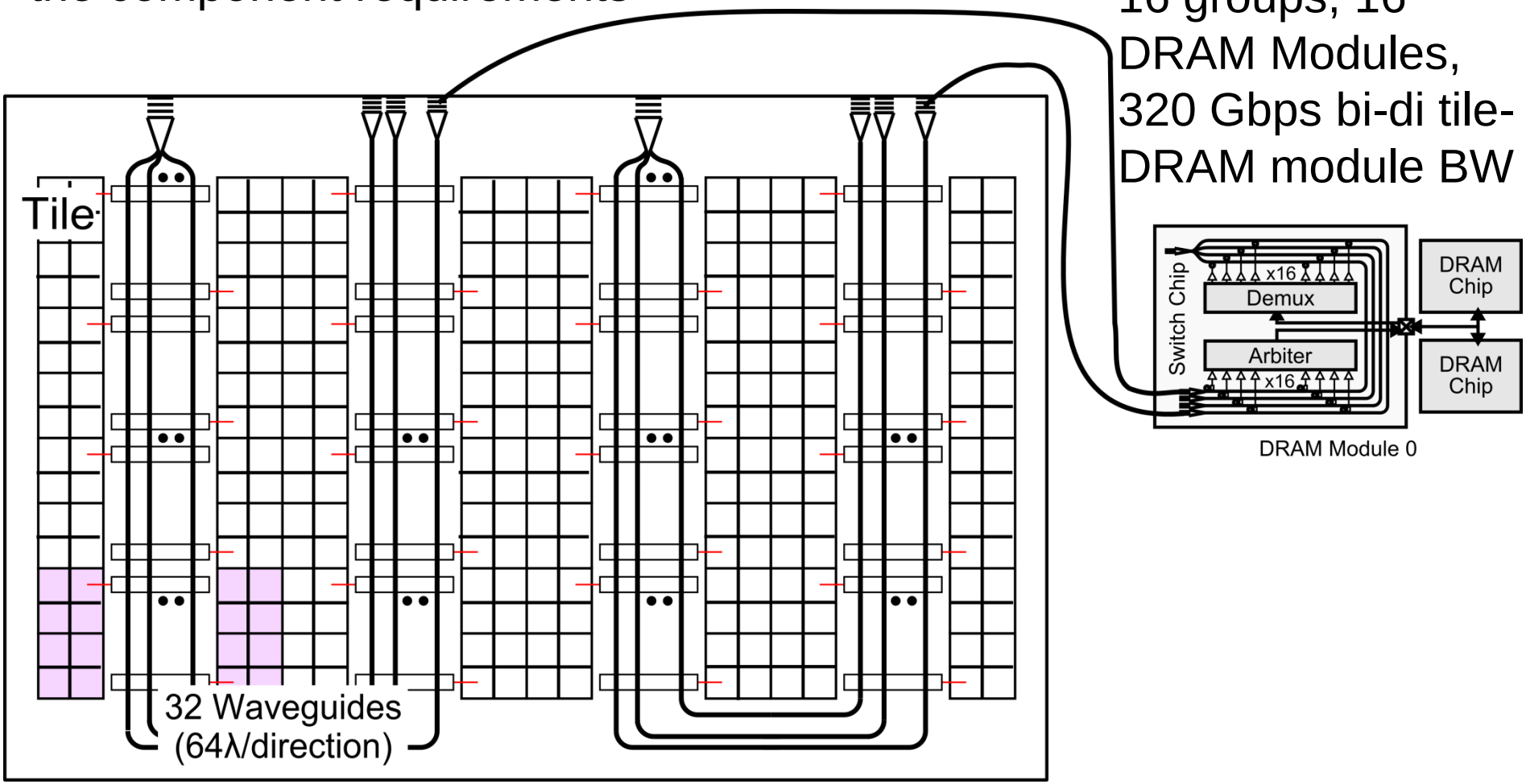
Ci = Core in Group i, DM = DRAM Module, S = Crossbar switch

- Shorten cross-chip electrical
- Photonic both part cross-chip and off-chip

Photonic LMGS: Physical Mapping

Network layout optimization significantly affects the component requirements

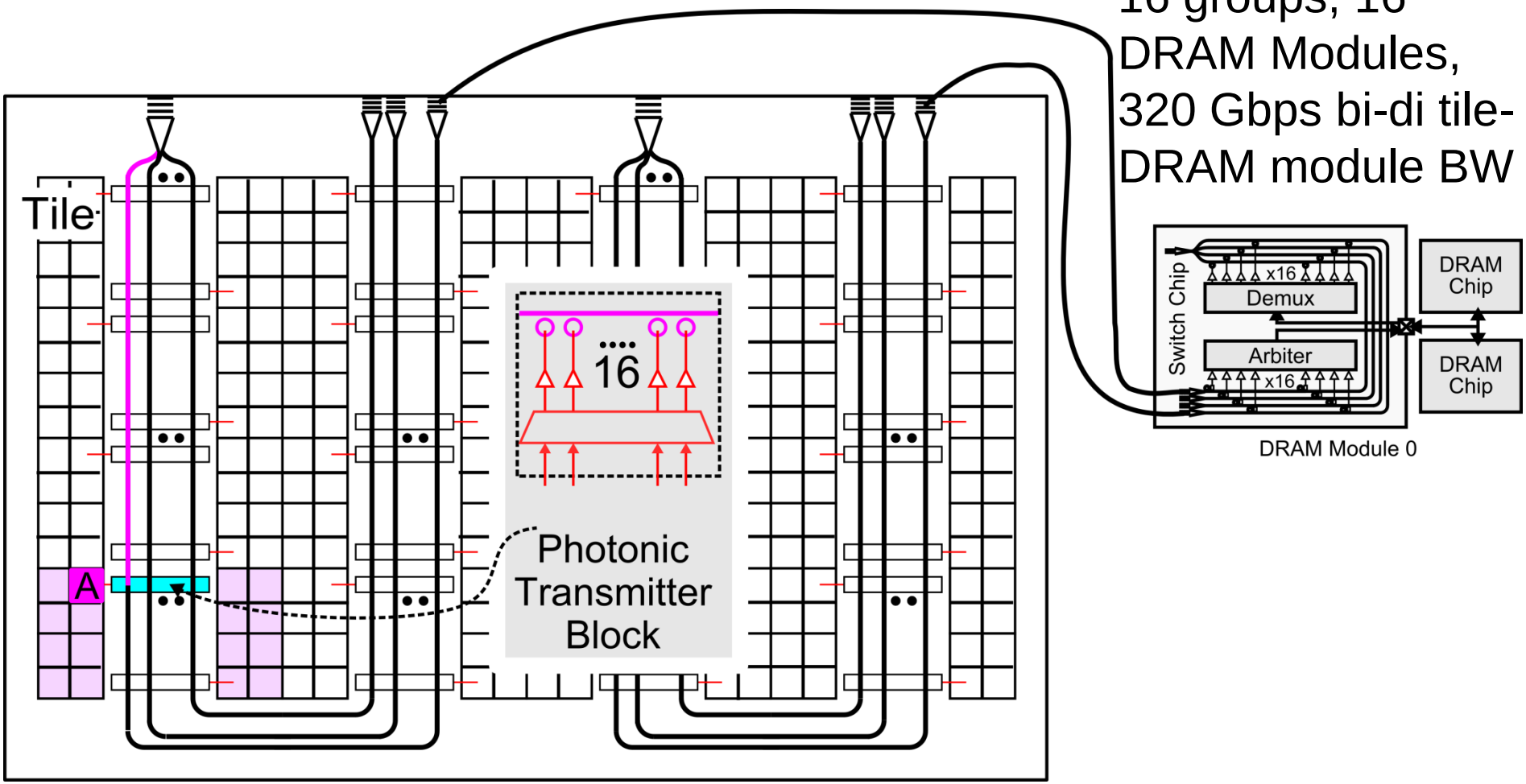
64-tile system w/ 16 groups, 16 DRAM Modules, 320 Gbps bi-di tile-DRAM module BW



[Joshi et al – PICA 2009]

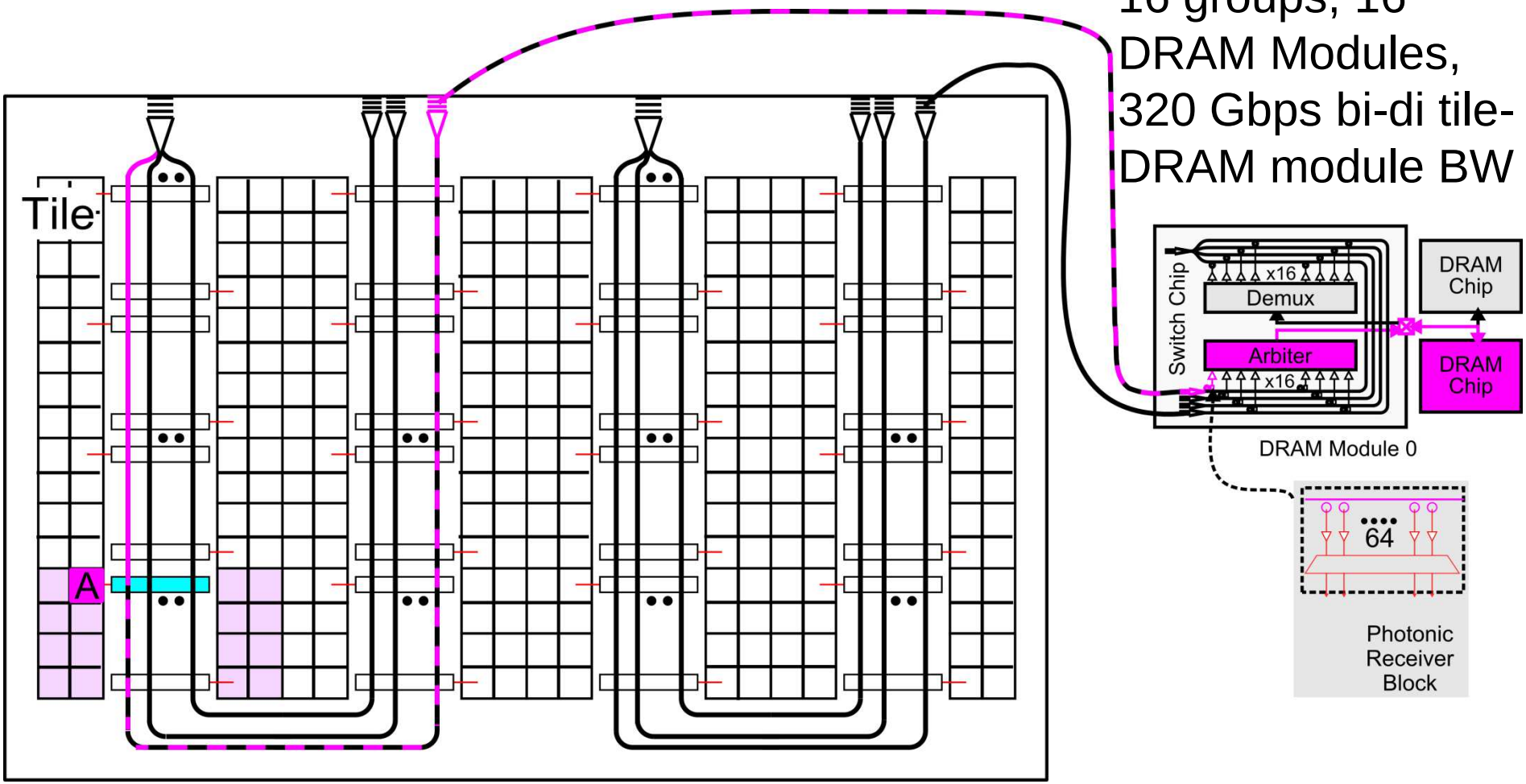
Photonic LMGS - U-shape

64-tile system w/ 16 groups, 16 DRAM Modules, 320 Gbps bi-di tile-DRAM module BW



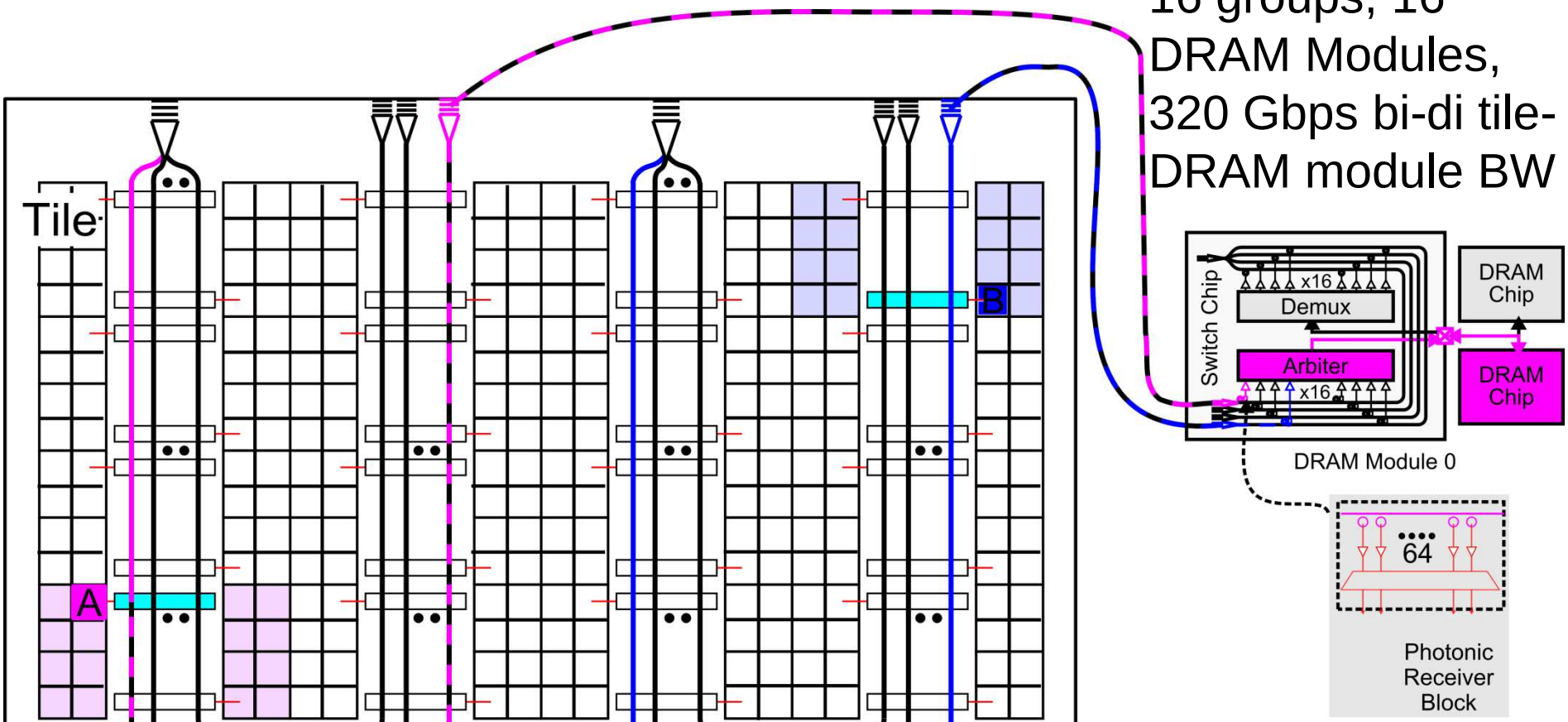
Photonic LMGS - U-shape

64-tile system w/ 16 groups, 16 DRAM Modules, 320 Gbps bi-di tile-DRAM module BW



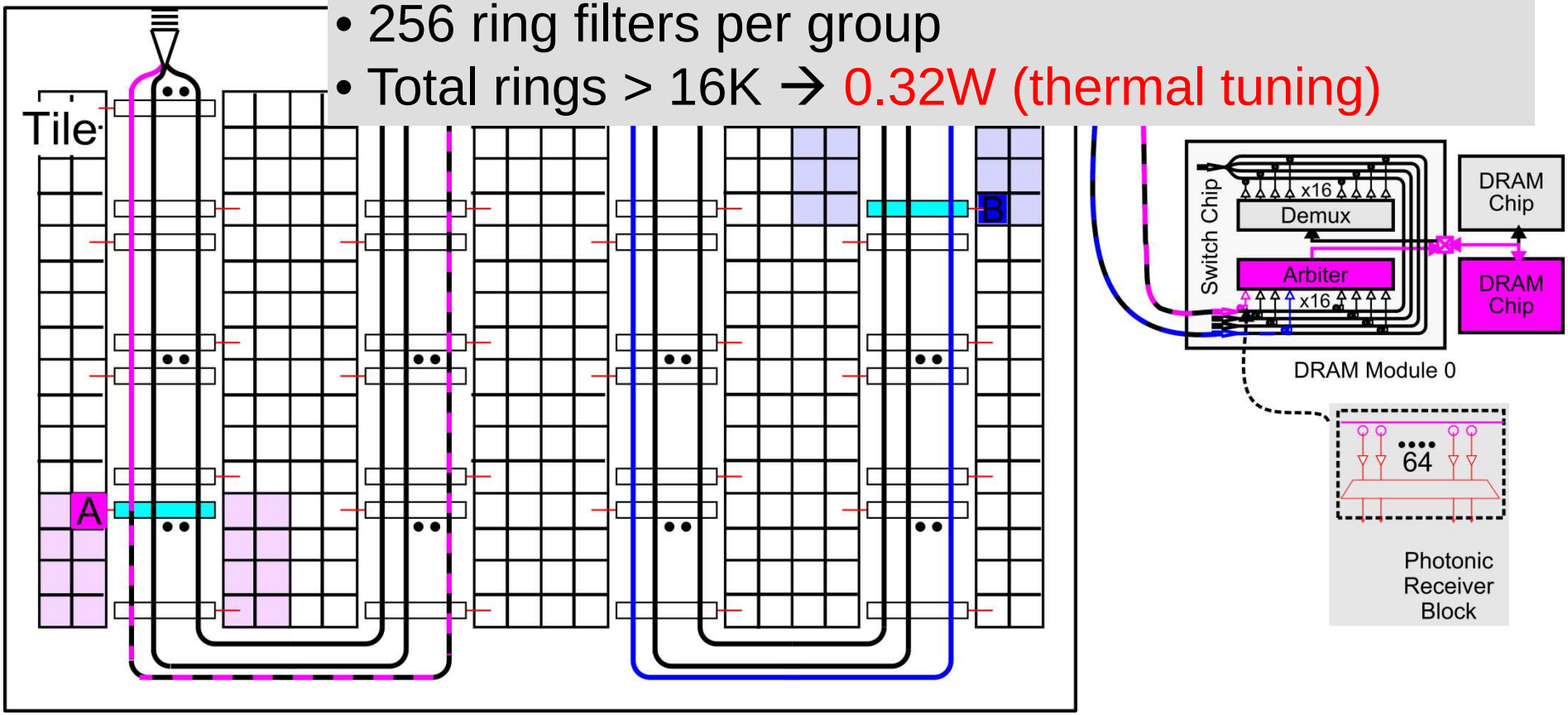
Photonic LMGS - U-shape

64-tile system w/ 16 groups, 16 DRAM Modules, 320 Gbps bi-di tile-DRAM module BW

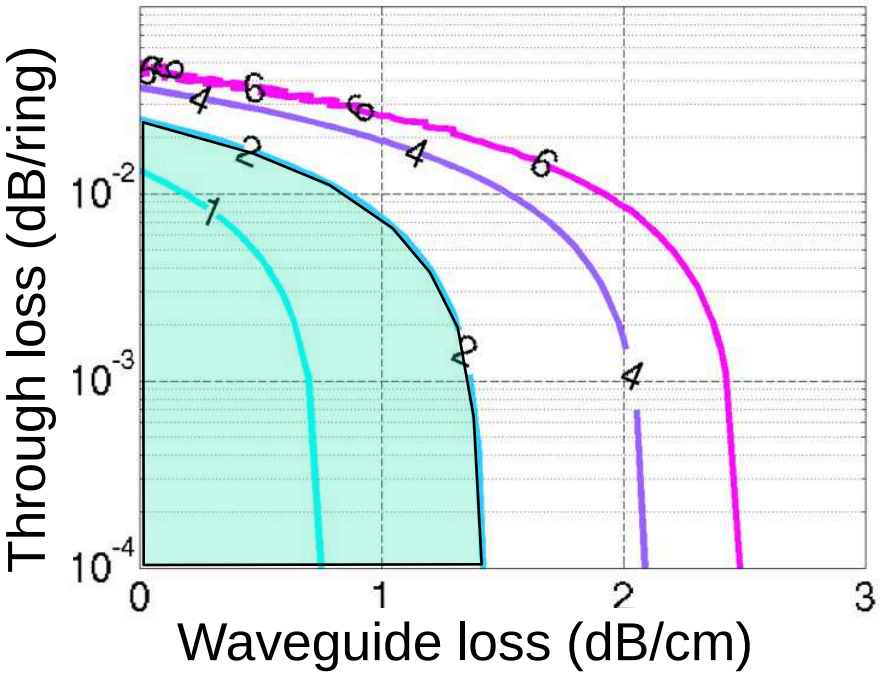


Photonic LMGS - U-shape

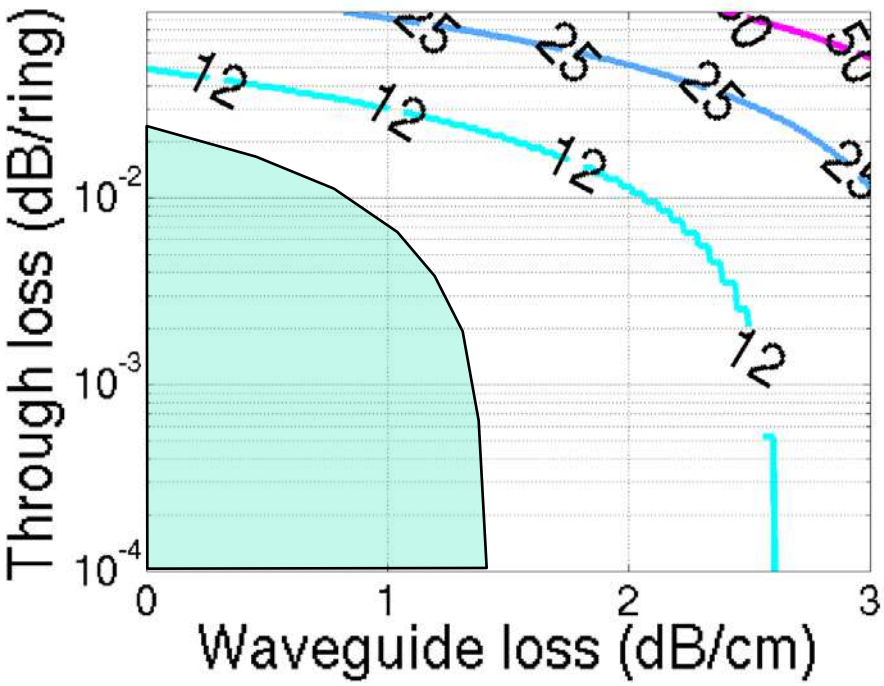
- 64 tiles
- 64 waveguides (for tile throughput = 128 b/cyc)
- 256 modulators per group
- 256 ring filters per group
- Total rings > 16K → 0.32W (thermal tuning)



Photonic device requirements in LMGS - U-shape



Optical Laser Power

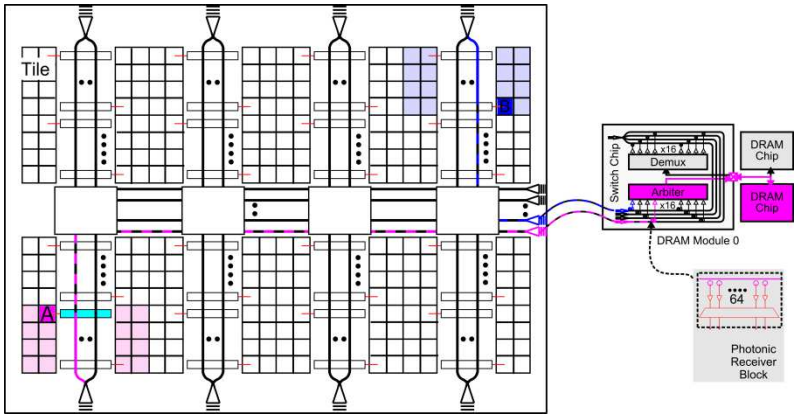


Die Area Overhead

Waveguide loss and Through loss limits for 2 W optical laser power

Photonic LMGS – ring matrix vs u-shape

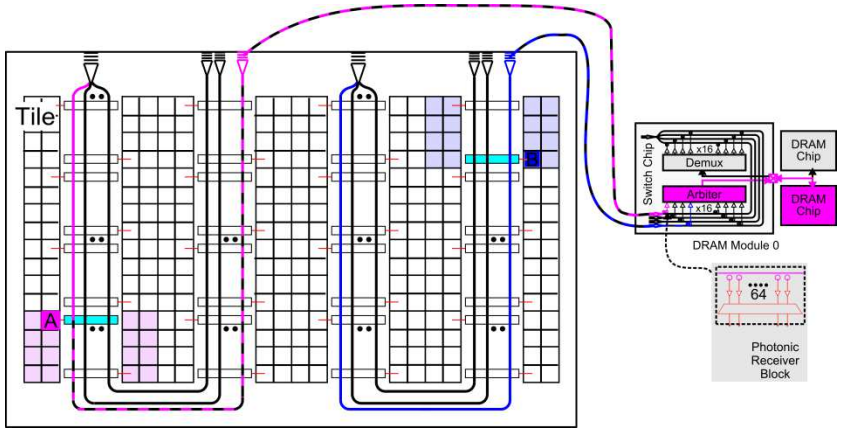
LMGS – ring matrix



- 0.64 W power for thermal tuning circuits
- 2 W optical laser power
  - Waveguide loss < 0.2 dB/cm
  - Through loss < 0.002 dB/ring

[Batten et al – Micro 2009]

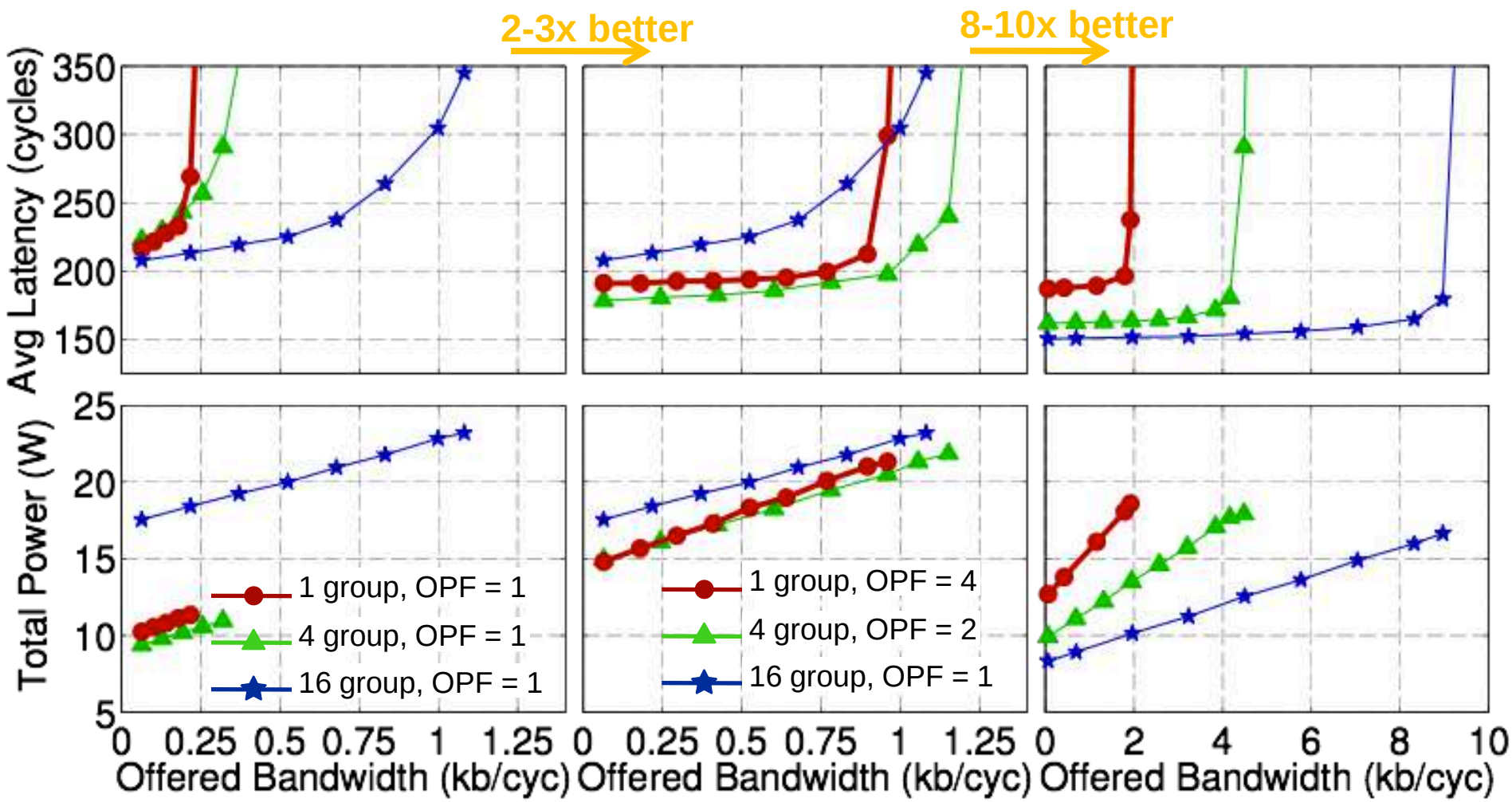
LMGS – u-shape



- 0.32 W power for thermal tuning circuits
- 2 W optical laser power
  - Waveguide loss < 1.5 dB/cm
  - Through loss < 0.02 dB/ring

[Joshi et al – PICA 2009]

Power-bandwidth tradeoff



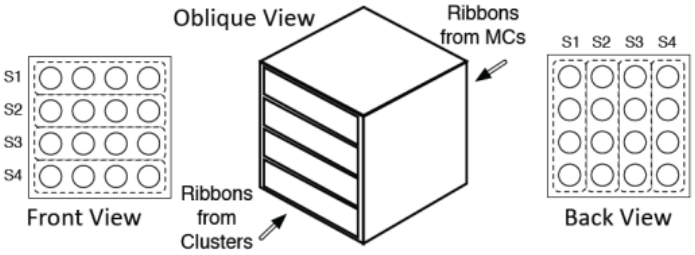
Electrical with grouping

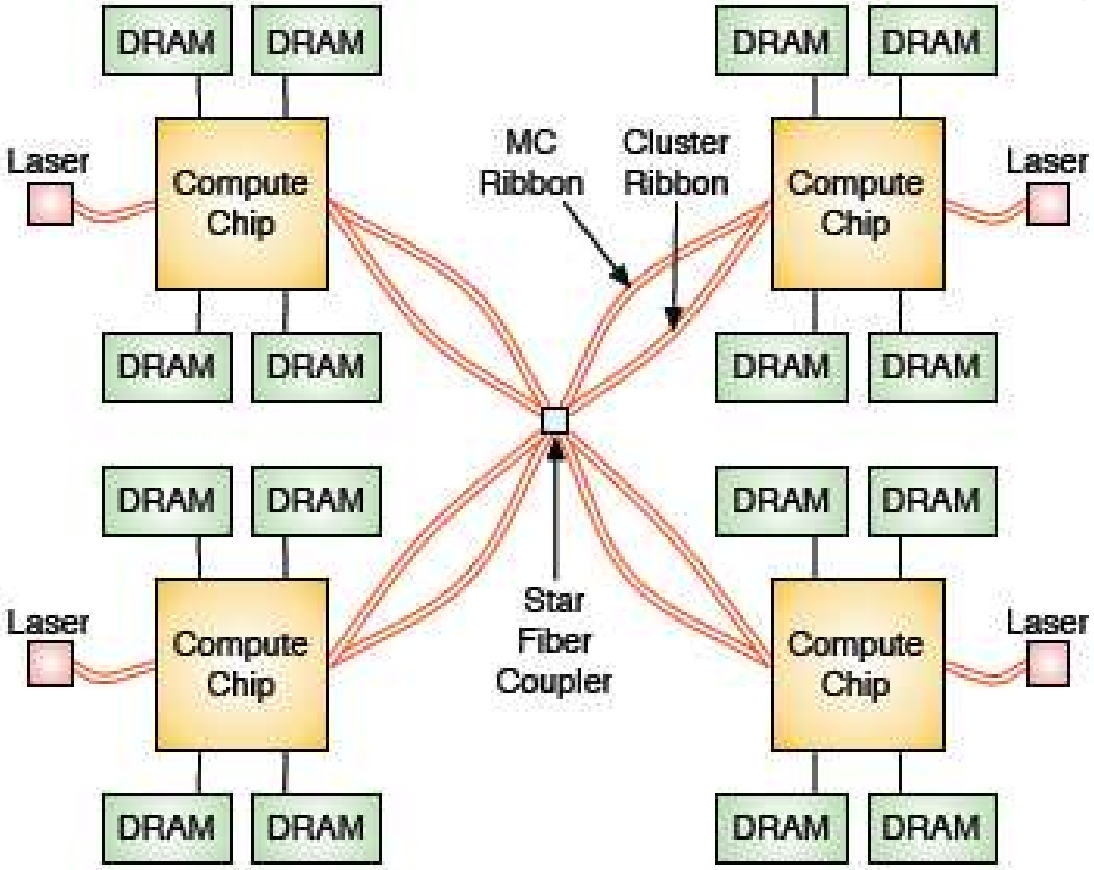
Electrical with grouping and over-provisioning

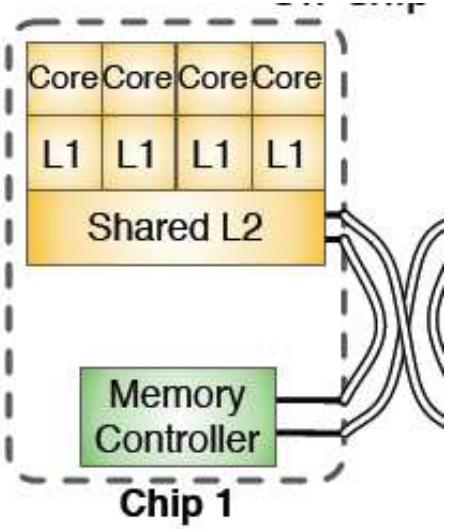
Optical with grouping and over-provisioning

System Organization – Defragmentation

[Beamer et al – ICS 2009]

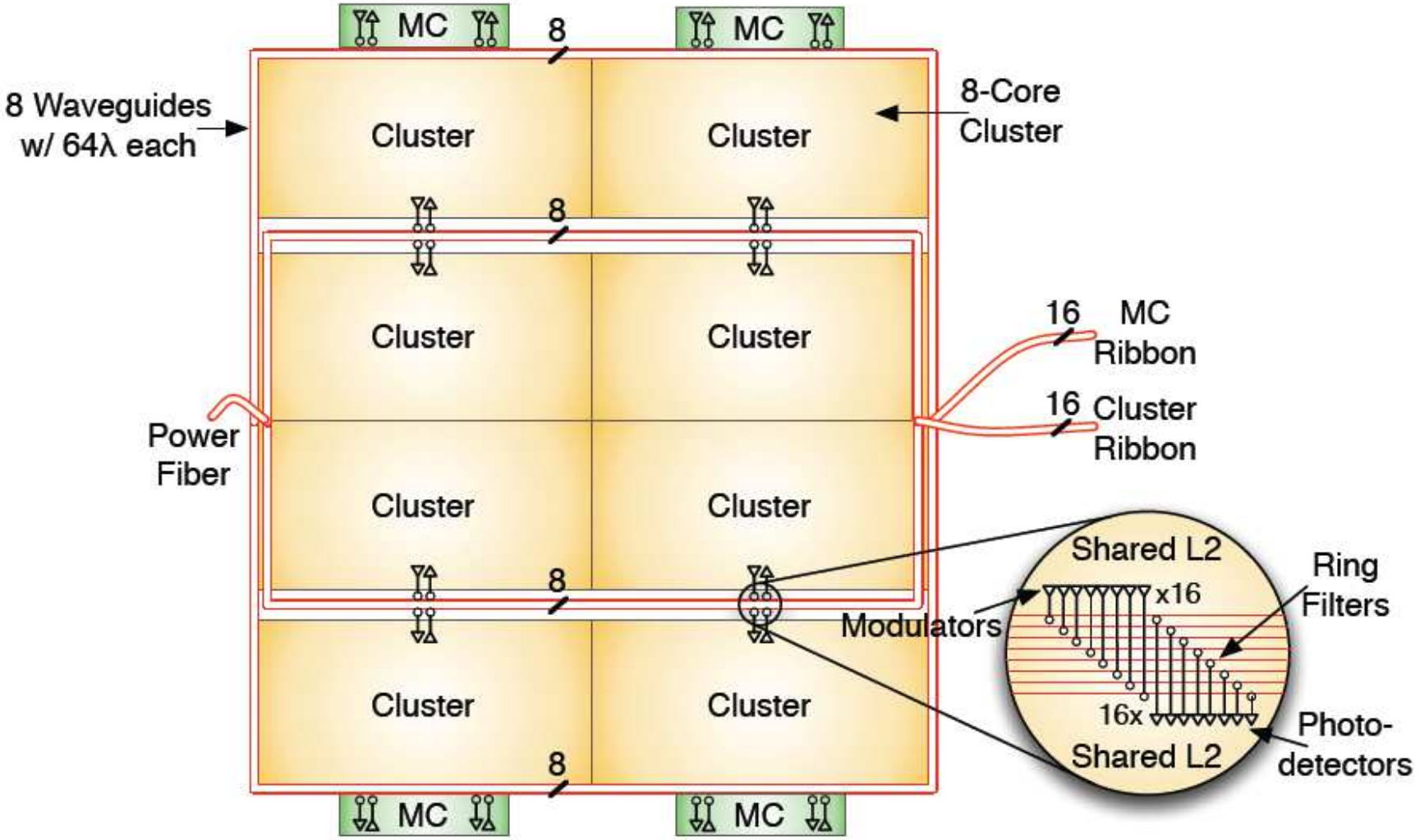






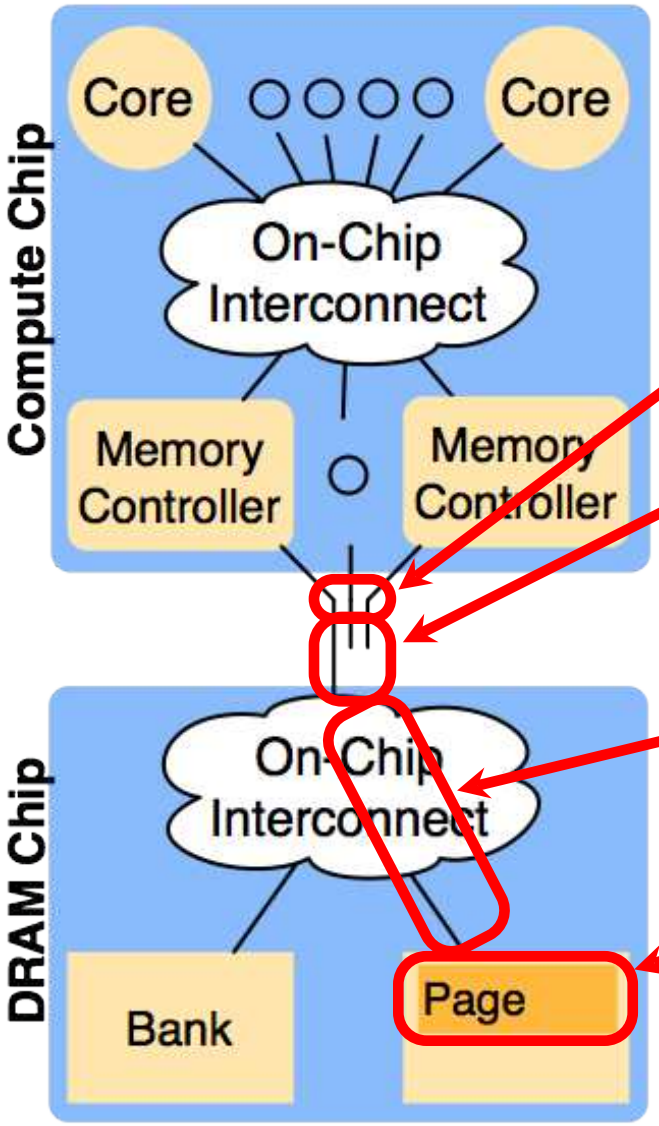
Example 256 core node – with 64 core dies

System Organization – Die view



64 core die supporting 256 core node

Electrical DRAM is also Limited



Pin-bandwidth on the compute chip

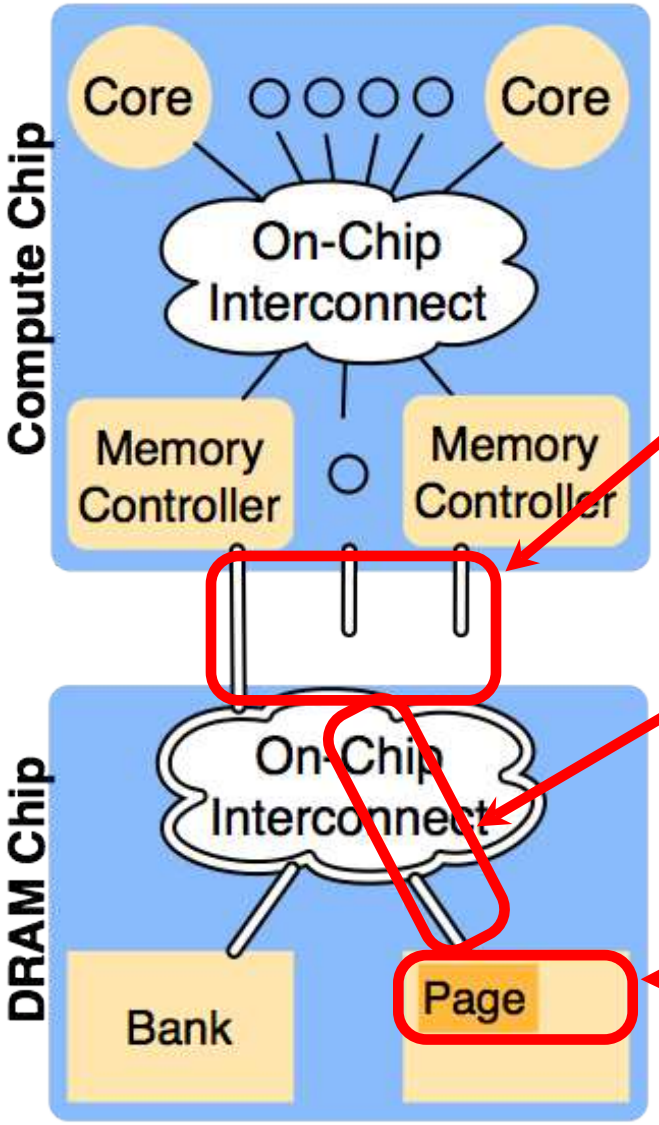
I/O energy to move between chips

Cross-chip energy within DRAM chip

Activation energy within DRAM chip

Solution: Silicon Photonics

[Beamer et al – ISCA 2010]



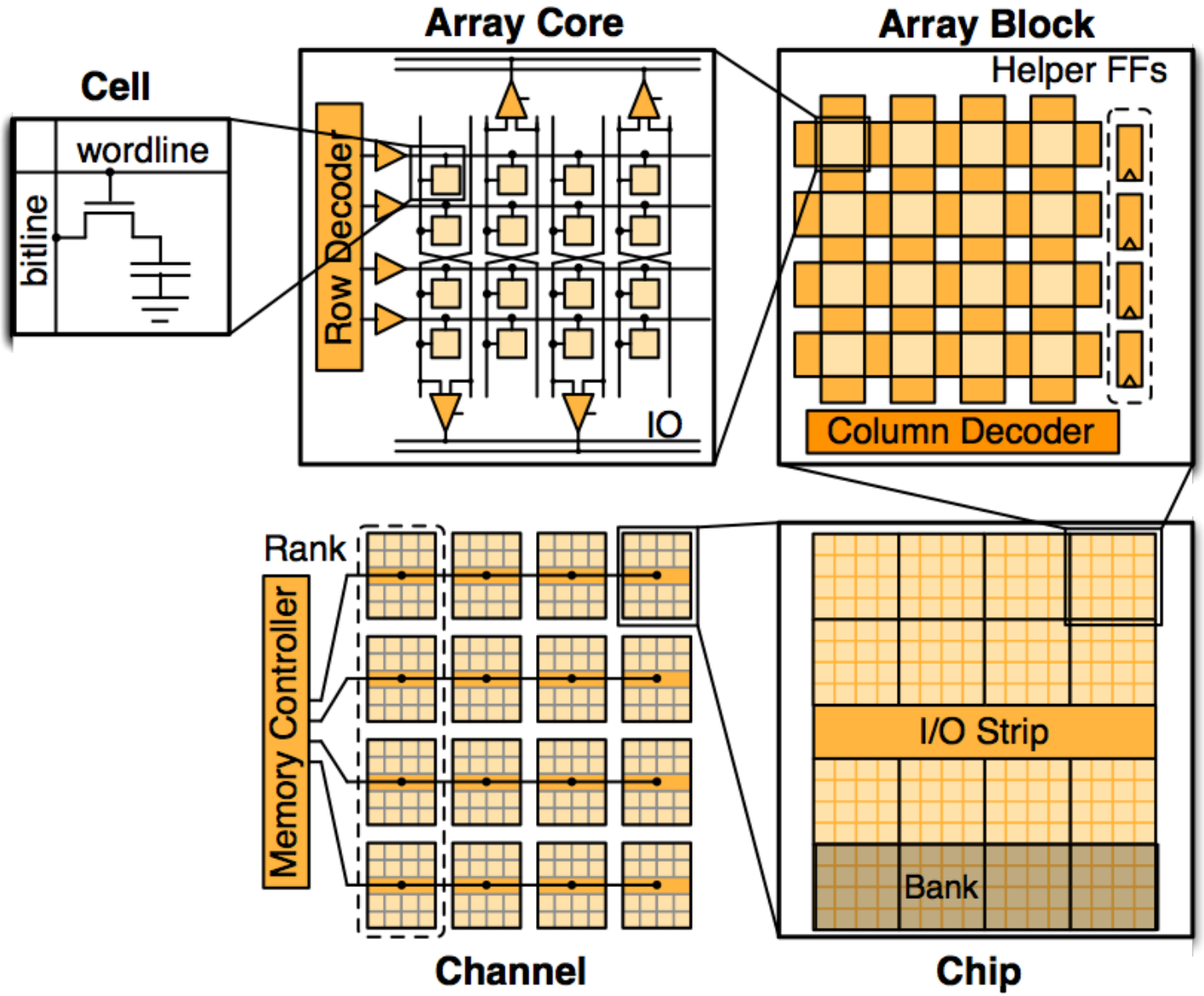
Great bandwidth density

Great off-chip energy efficiency

Costs little additional energy to use on-chip after off-chip

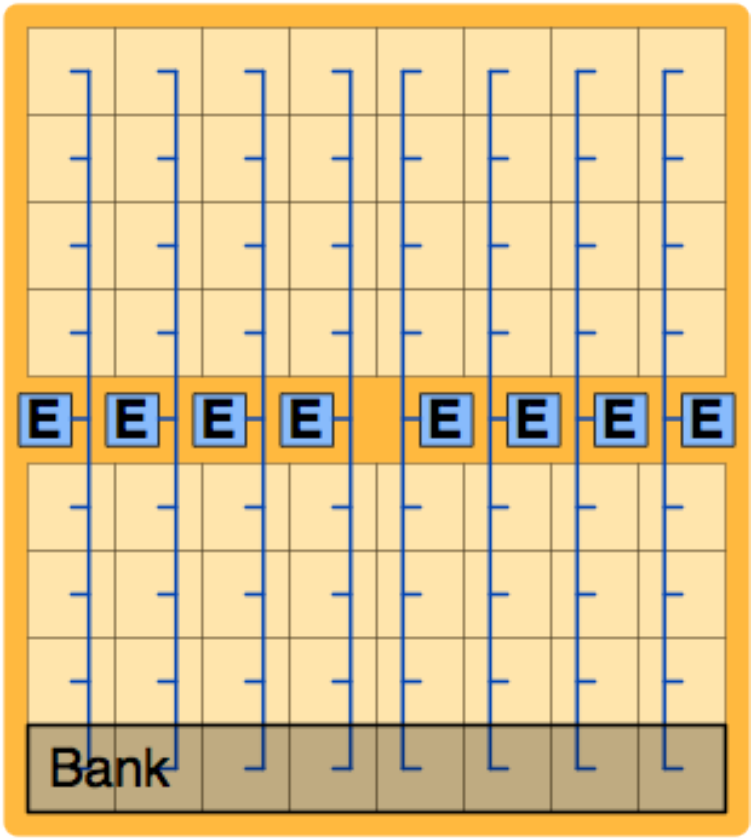
Enables page size reduction

Current DRAM Structure



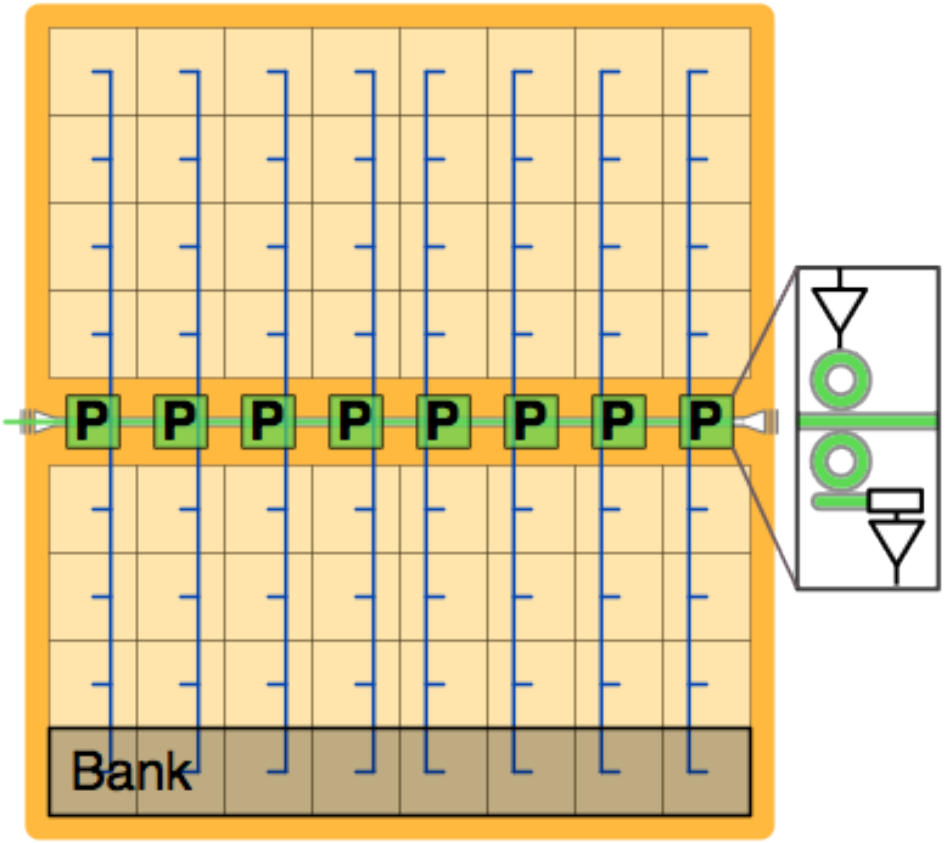
Photonics to the Chip

Electrical Baseline (E1)



E Electrical Off-Chip Driver

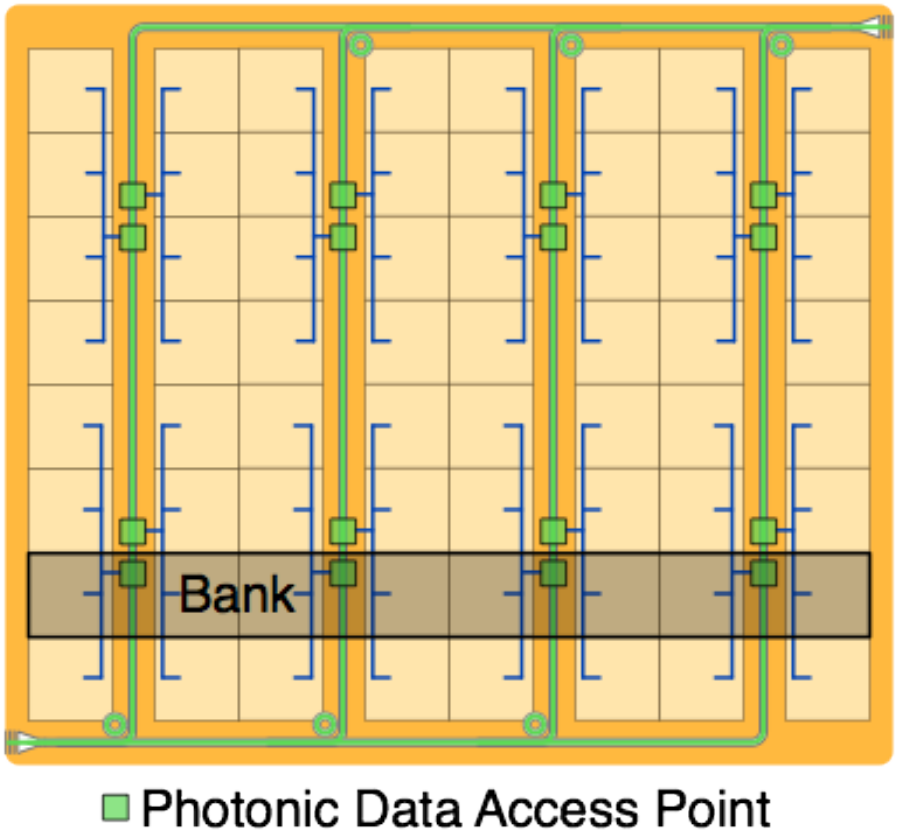
Photonics Off-Chip w/Electrical On-Chip (P1)



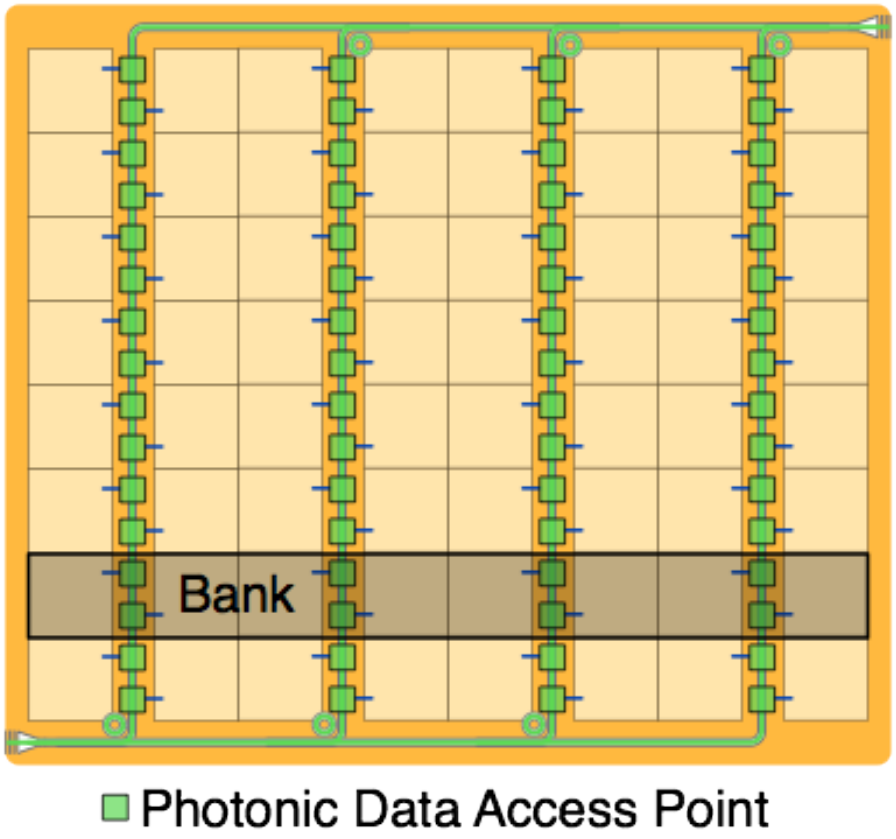
P Photonic Data Access Point

Photonics Into the Chip

2 Data Access Points per Column (P2)

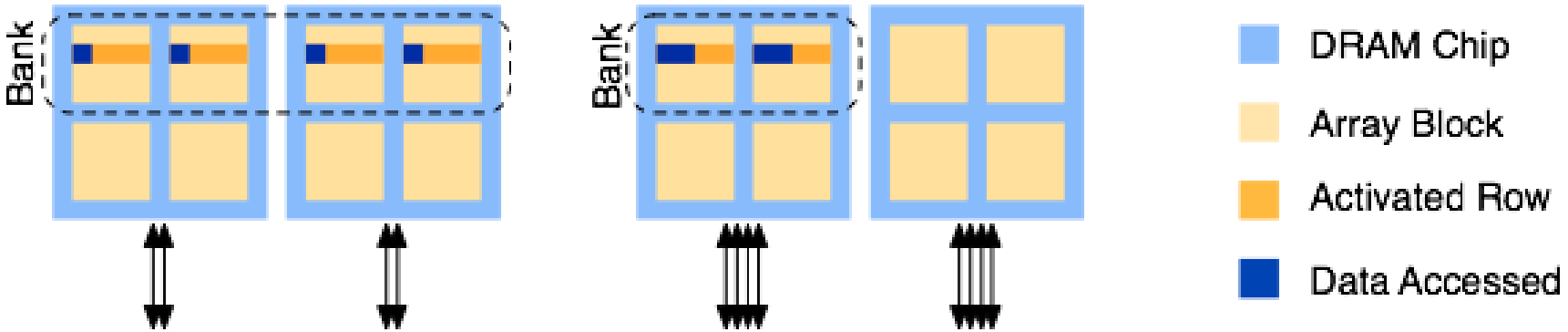


8 Data Access Points per Column (P8)



Reducing Activate Energy

- Want to activate less bits while achieving the same access width
- Increase number of I/Os per array core, which decreases page size
  - Compensate the area hit by smaller photonic off-chip I/O



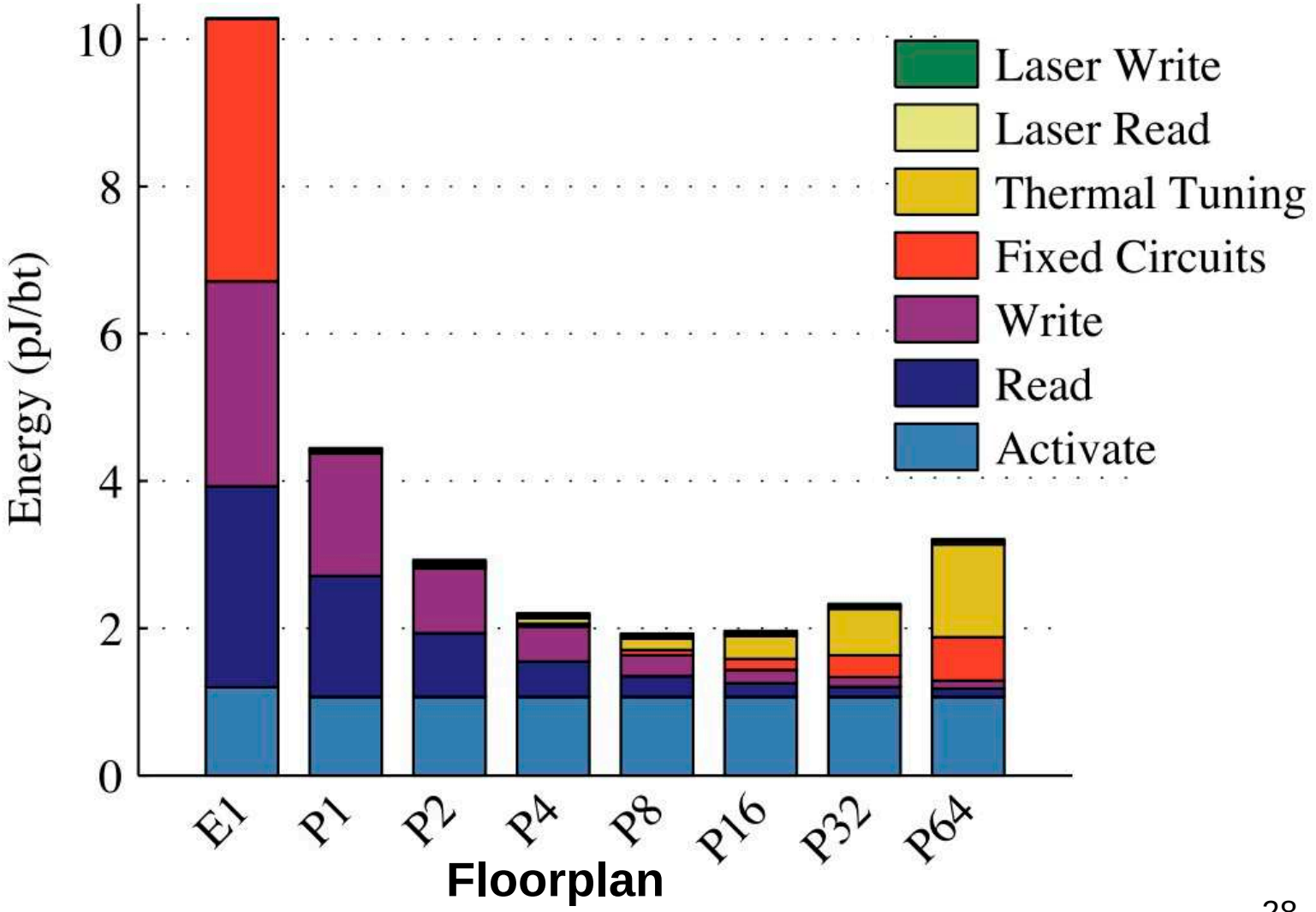
Initial Design

Double the I/Os (and bandwidth)

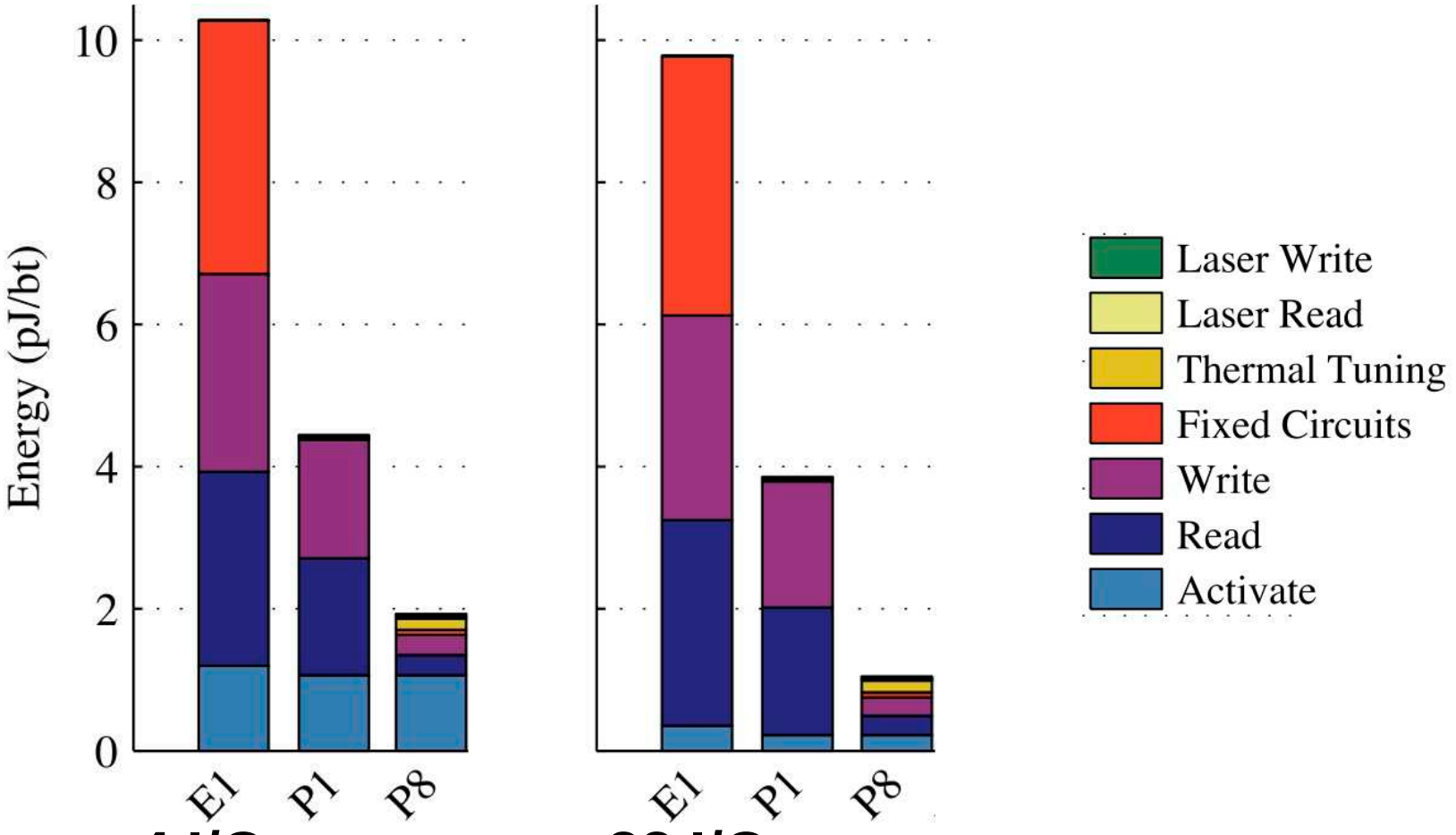
Methodology

- Photonic Model - aggressive and conservative projections
- DRAM Model - Heavily modified CACTI-D
- Custom C++ architectural simulator running random traffic to animate models
- Setup is configurable, in this presentation:
  - 1 chip to obtain 1GB capacity with >500Gbps of bandwidth provided by 64 banks

Energy for On/Off-Chip



Reducing Row Size



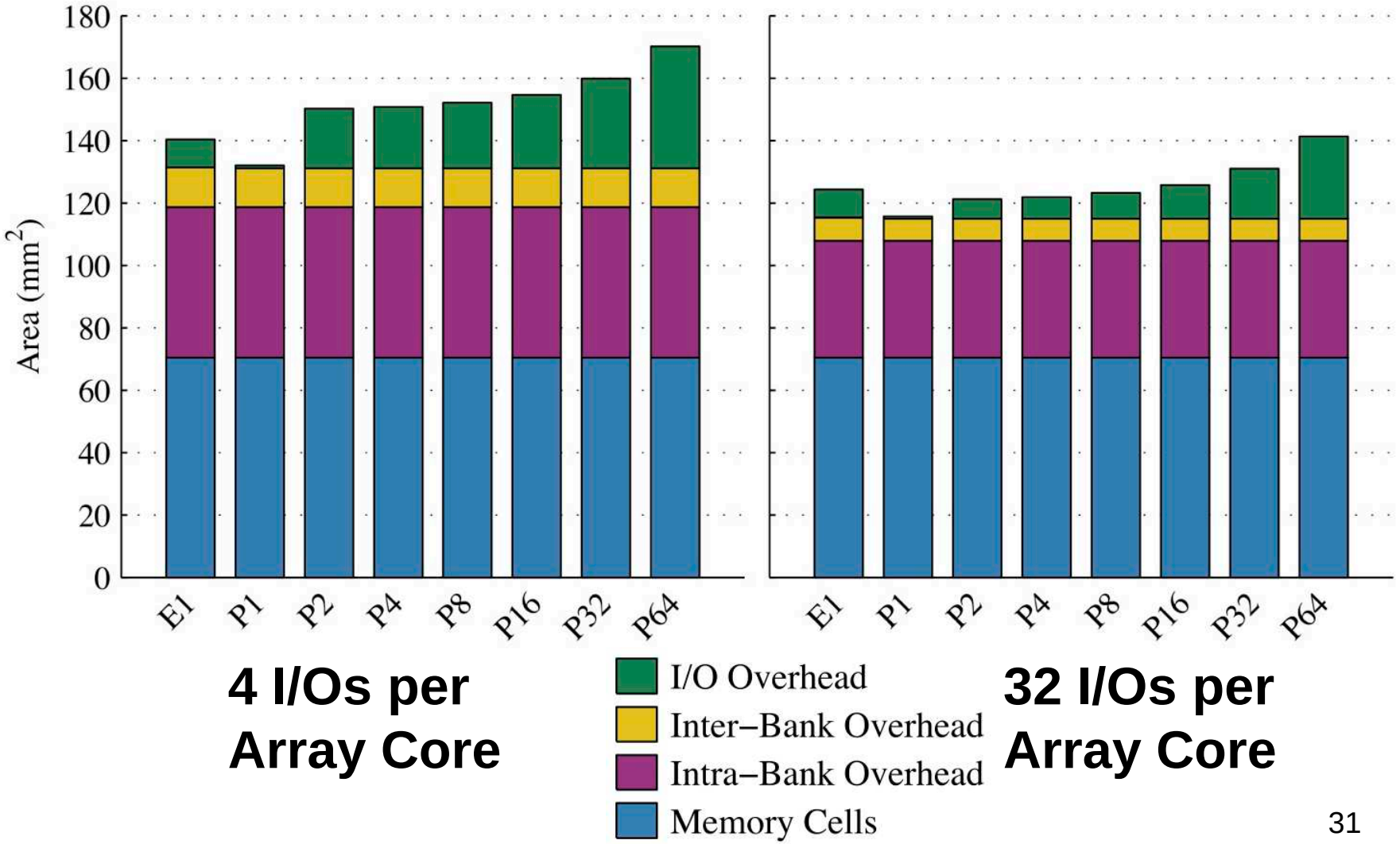
4 I/Os per Array Core

32 I/Os per Array Core

Latency Not a Big Win

- Latency marginally better
- Most of latency is within array core
- Since array core mostly unchanged, latency only slightly improved by reduced serialization latency

Area Neutral

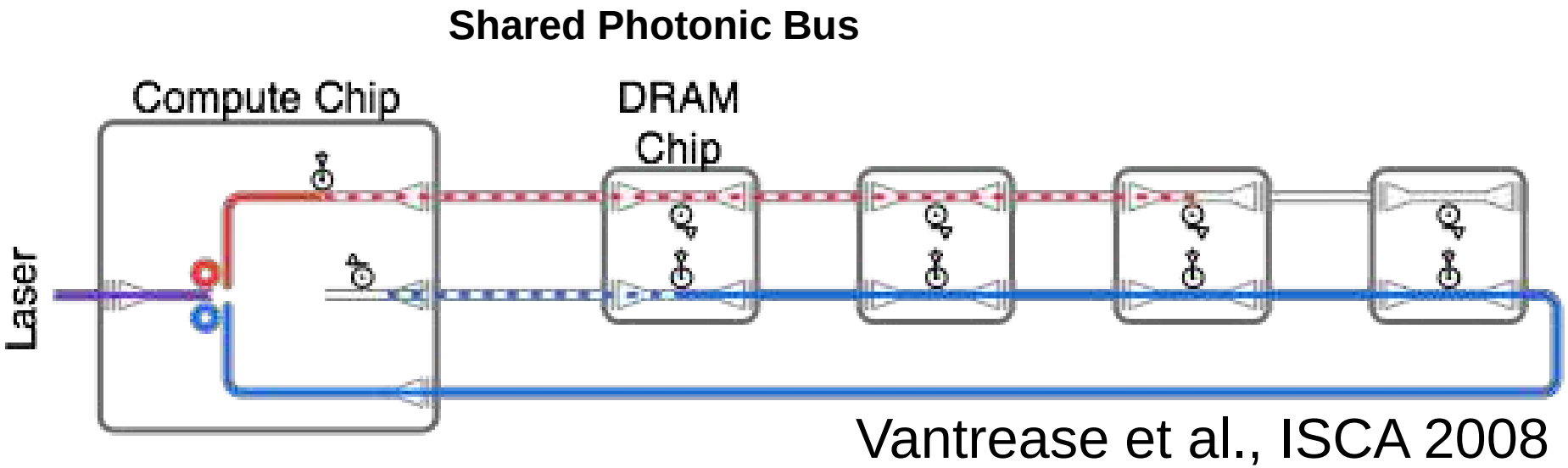


4 I/Os per Array Core

32 I/Os per Array Core

Scaling Capacity

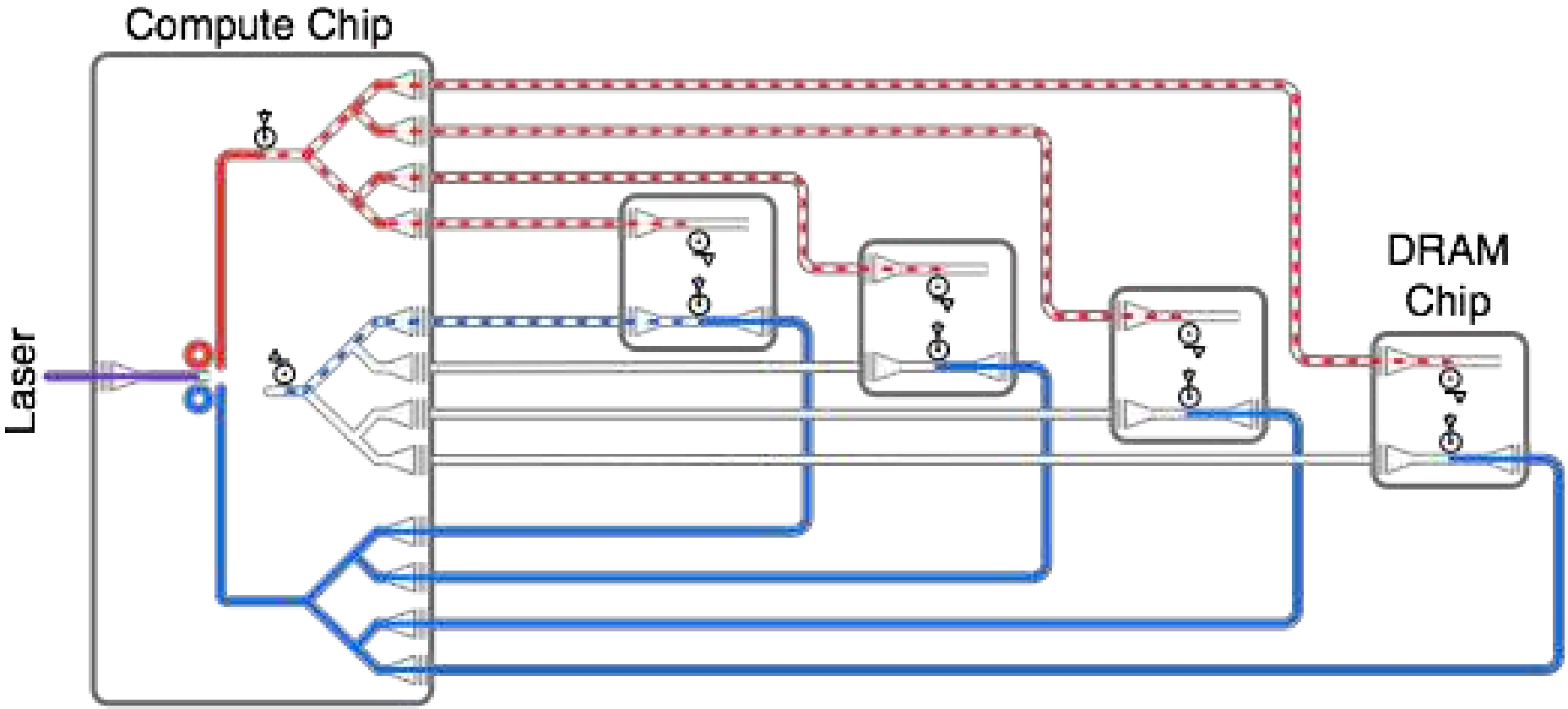
- ***Motivation:*** allow the system to increase capacity without increasing bandwidth



Vantrease et al., ISCA 2008

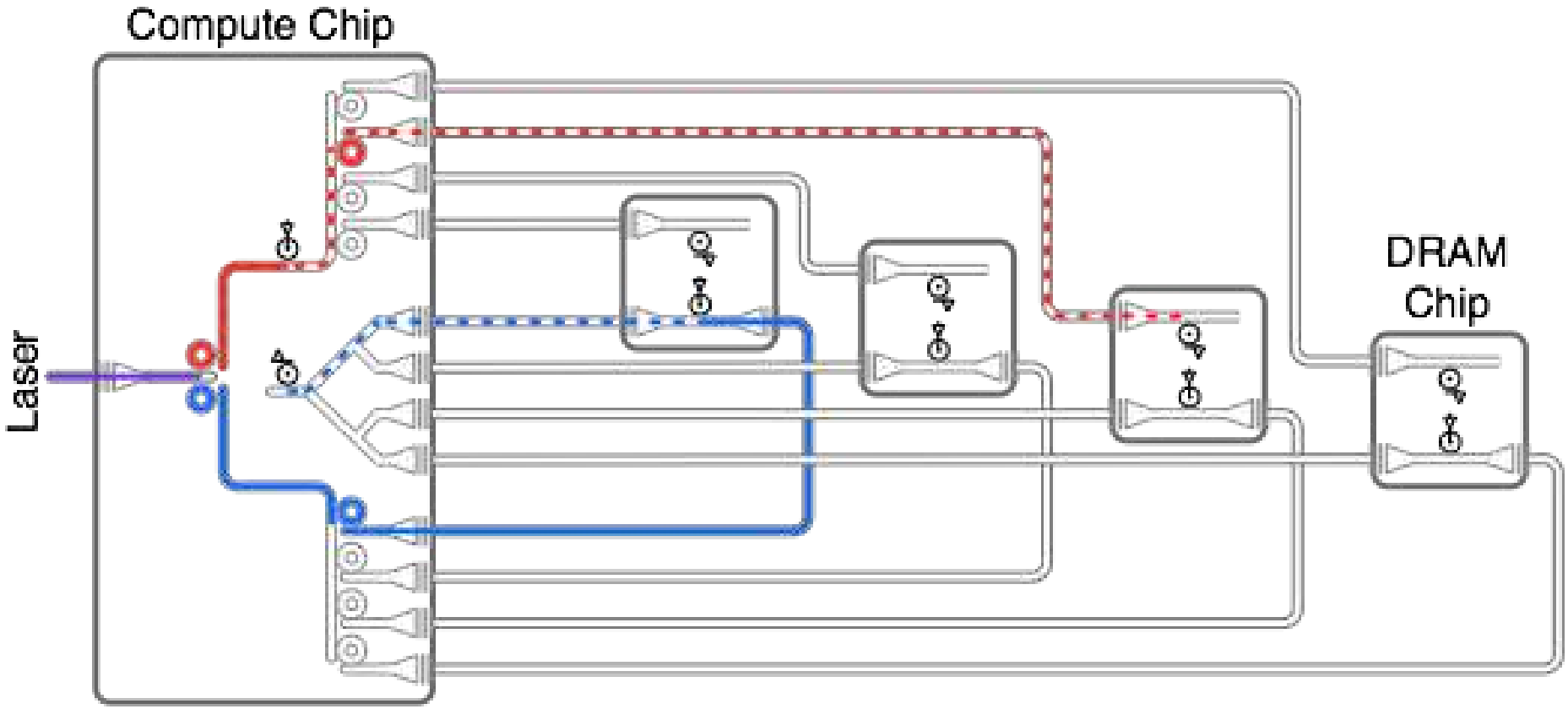
- ***Disadvantage*****:** high path loss (grows exponentially) due to couplers and waveguide

Split Photonic Bus



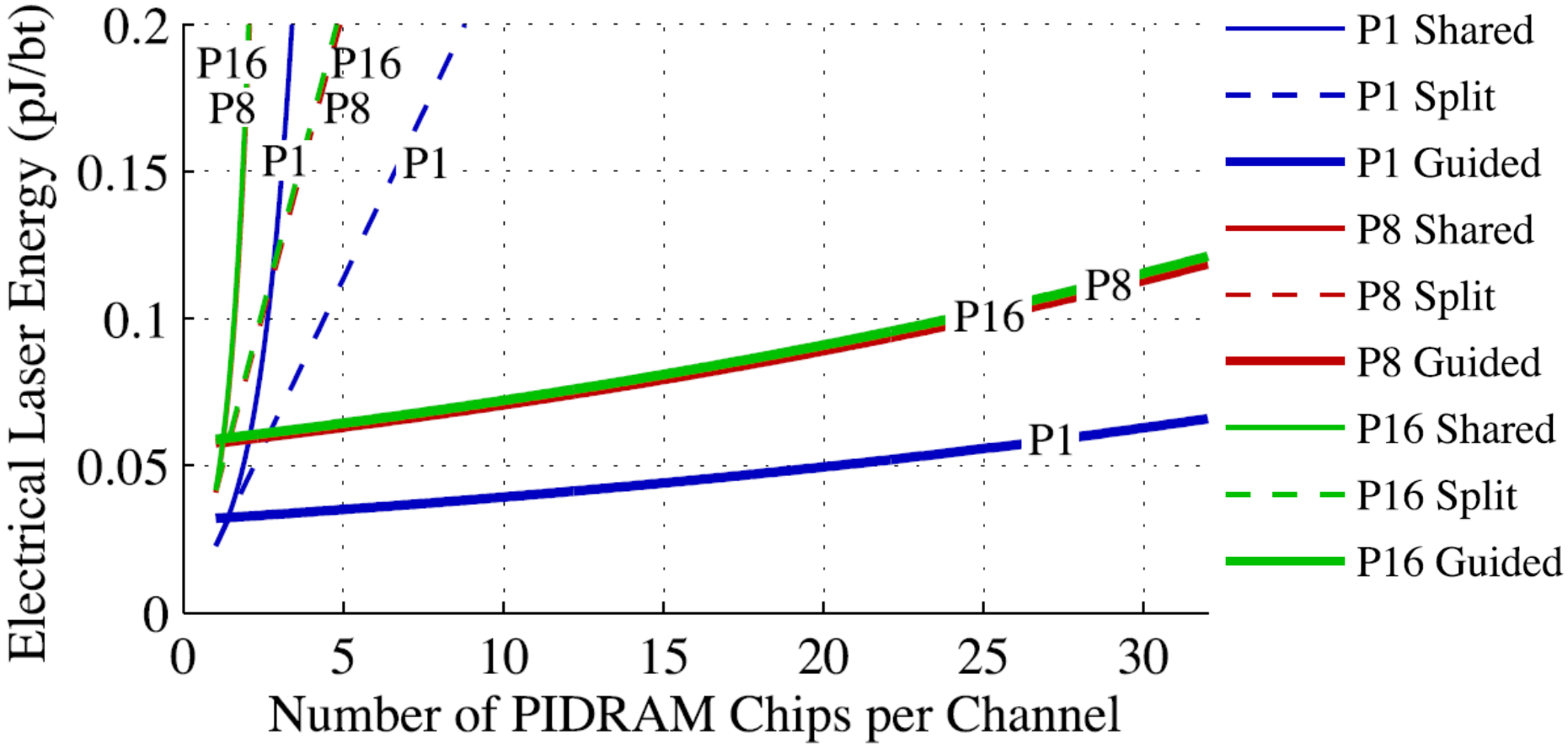
- ***Advantage:*** much lower path loss
- ***Disadvantage:*** all paths lit

Guided Photonic Bus



- ***Advantage:*** only 1 low loss path lit

Scaling Results



Aggressive Photonic Device Specs

With Photonics...

- 10x memory bandwidth for same power
- Higher memory capacity without sacrificing bandwidth
- Area neutral
- Easily adapted to other storage technologies

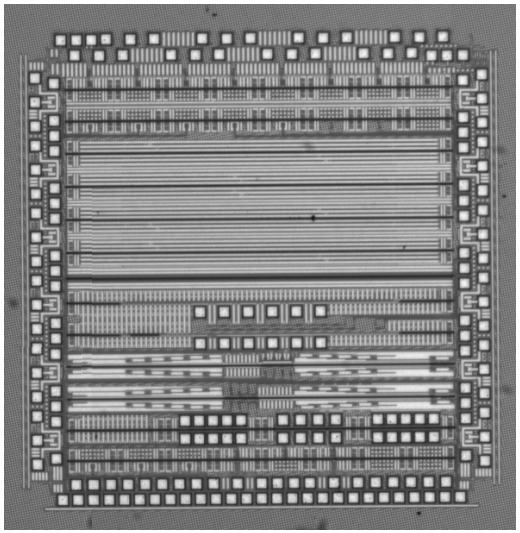
Conclusion

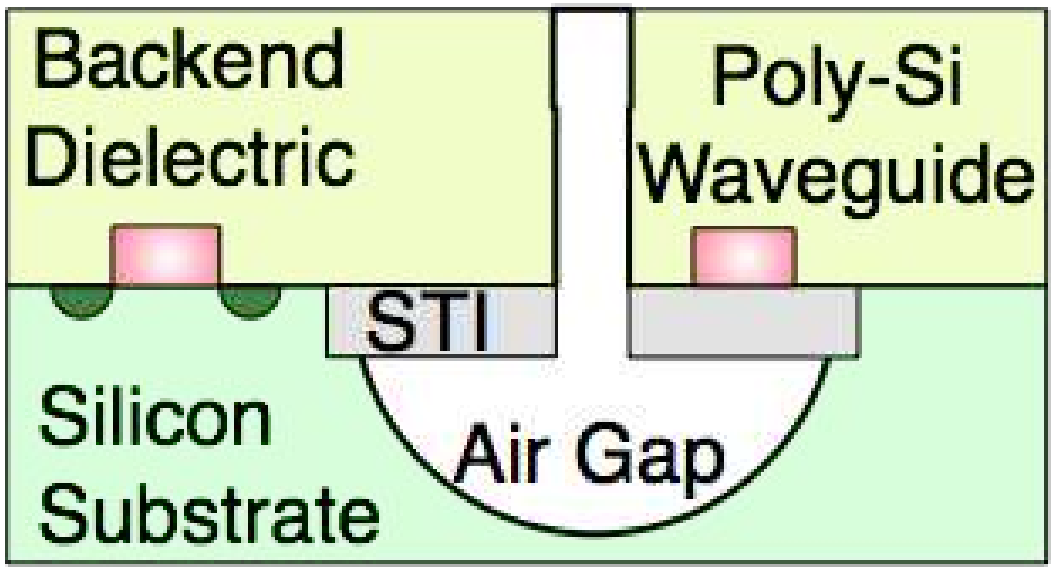
- Computer interconnects are very complex micro-communication systems
- Cross-layer design approach is needed to solve the on-chip and off-chip interconnect problem
  - Most important metrics
    - Bandwidth-density (Gb/s/um)
    - Energy-efficiency (mW/Gb/s)
  - Monolithic CMOS-photonics can improve the throughput by 10-20x
  - But, need to be careful
    - Optimize network design (electrical switching, optical transport)
    - Use aggregation to increase link utilizations
    - Optimize physical mapping (layout) for low optical insertion loss

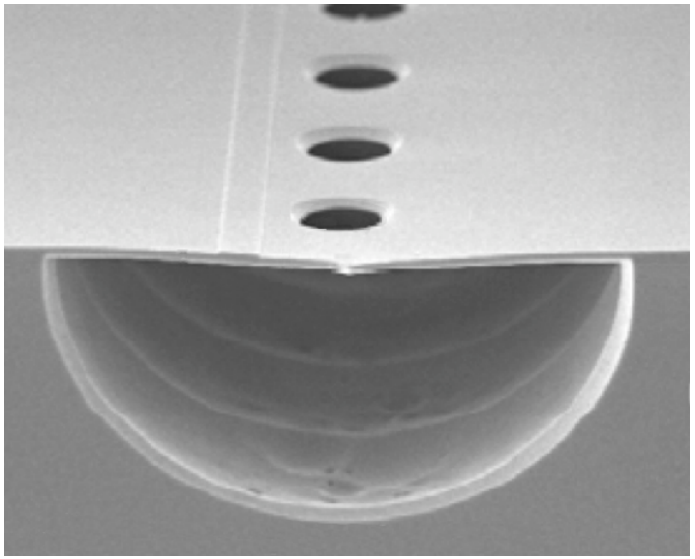
Backup Slides

Photonic Technology

- Monolithically integrated silicon photonics being researched by MIT Center for Integrated Photonic Systems (CIPS)





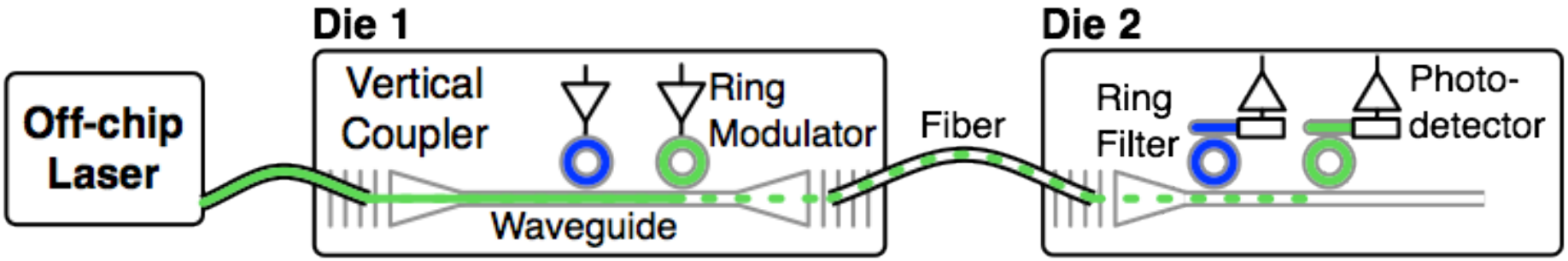


Orcutt et al., CLEO 2008

Holzwarth et al., CLEO 2008

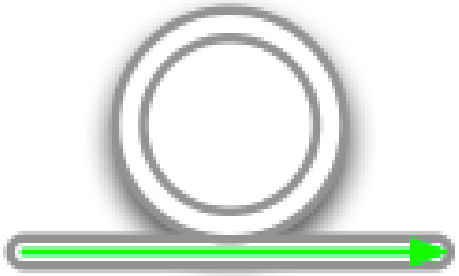
Photonic Link

- Each wavelength can transmit at 10Gbps
- Dense Wave Division Multiplexing (DWDM)
  - 64 wavelengths per direction in same media

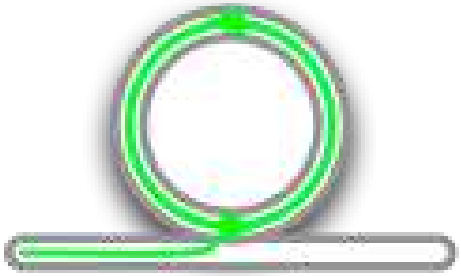


Rough Comparison	Electrical	Photonic
Off-Chip I/O Energy (pJ/bit)	5	0.150
Off-Chip BW Density (Tbps/mm^2)	1.5	50.000

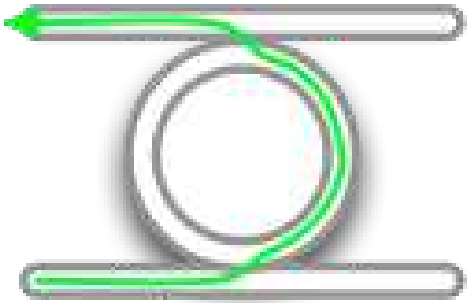
Resonant Rings



light not resonant



resonant light

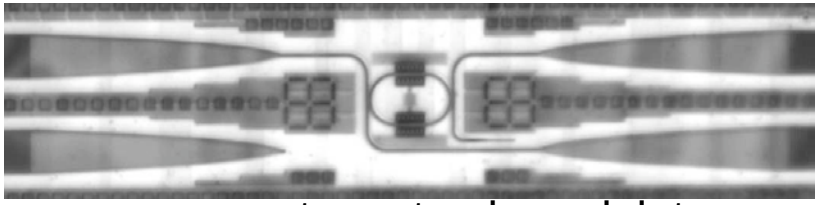


resonant light w/ drop path

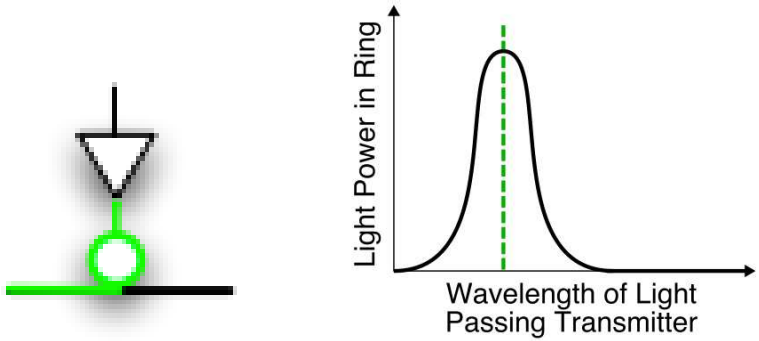
figures inspired by [Vantrease, ISCA ’08]

Ring Modulators

- **Modulator** uses charge injection to change resonant wavelength
- When resonant light passes it mostly gets trapped in ring



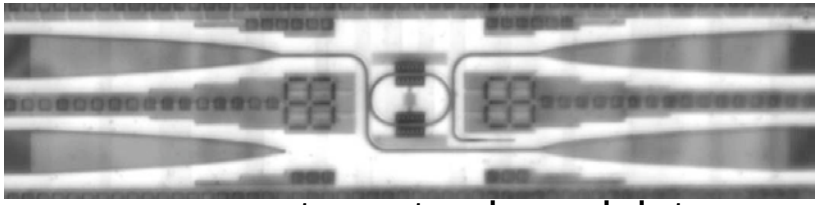
resonant racetrack modulator



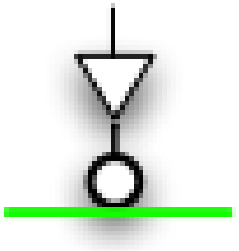
modulator off

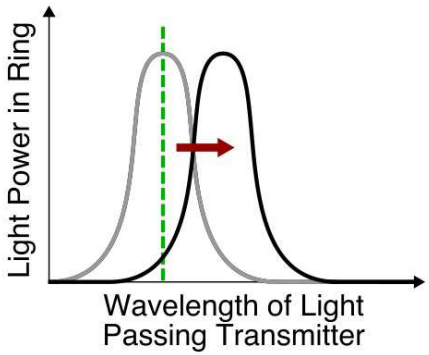
Ring Modulators

- **Modulator** uses charge injection to change resonant wavelength
- When resonant light passes it mostly gets trapped in ring



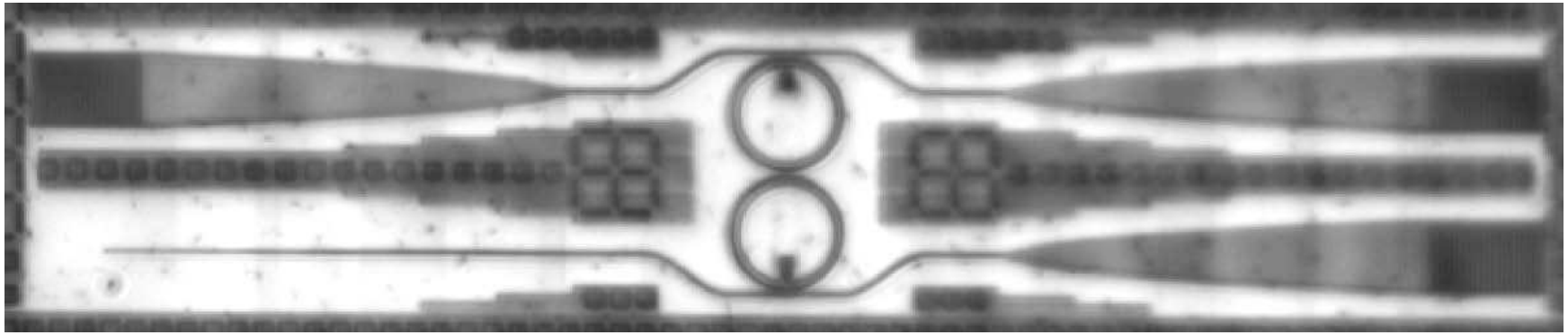
resonant racetrack modulator

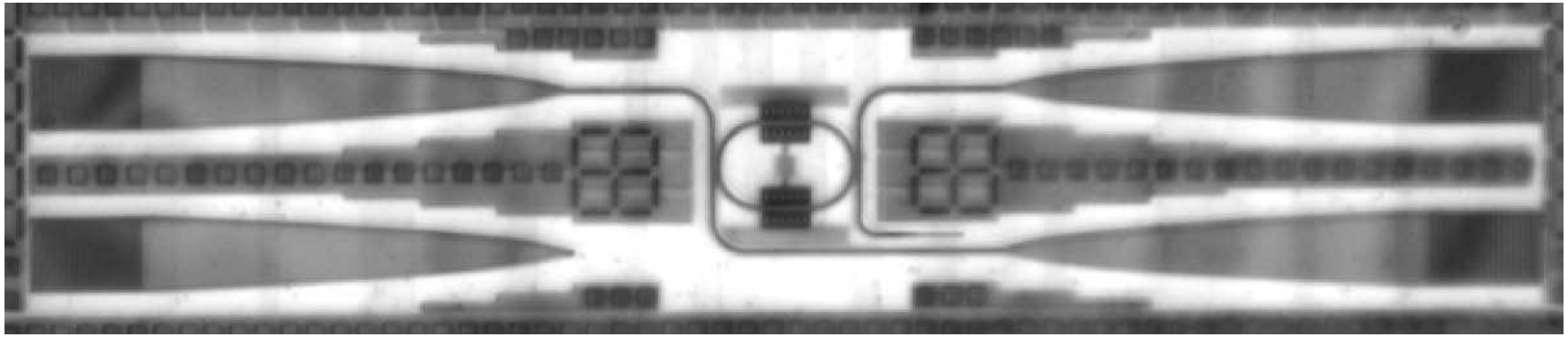




modulator on

Photonic Components

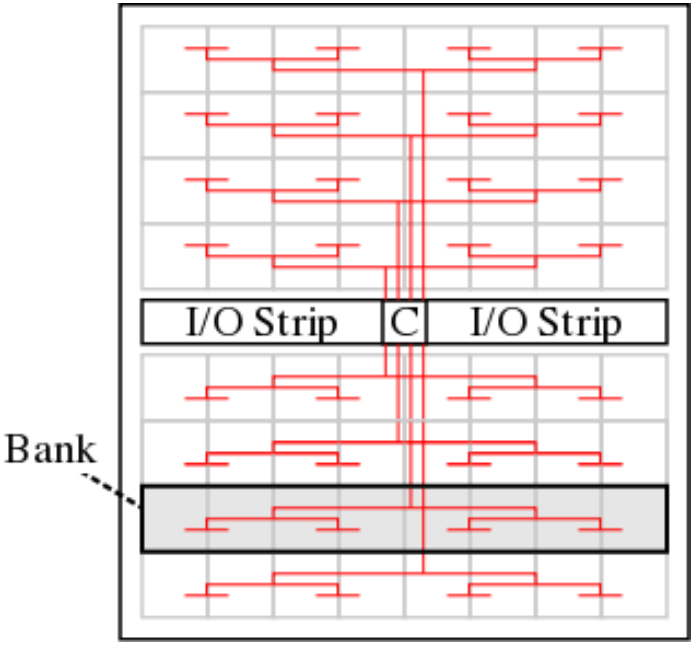




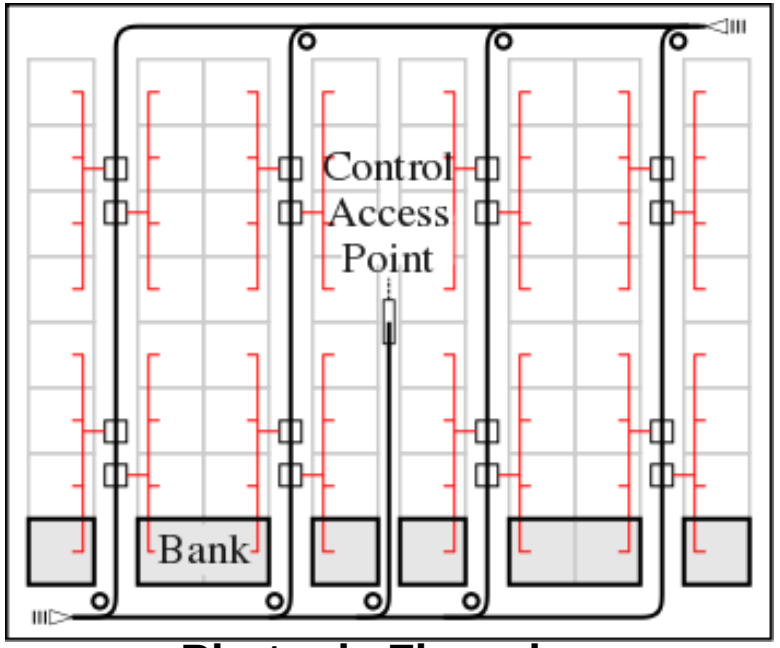
Why 5pJ/b for Electrical?

- Prior work has claimed lower than our forecasted 5pJ/b for off-chip electrical I/O
  - 2.24 pJ/b @ 6.25Gbps (Palmer et al., ISSCC 2007)
  - 1.4 pJ/b @ 10Gbps (O'Mahony et al., ISSCC 2010)
- Some important differences to consider:
  - We assume 20Gbps per pin
    - Otherwise will definitely be pin limited
    - At higher data rates it is hard to be as energy efficient: 8-13pJ/b @ 16Gbps (Lee et al., JSSC 2009)
- DRAM process has slower transistors leading to less energy efficient drivers
- Background energy averaged in (clocking, fixed energy, not 100% utilization)

Control Distribution



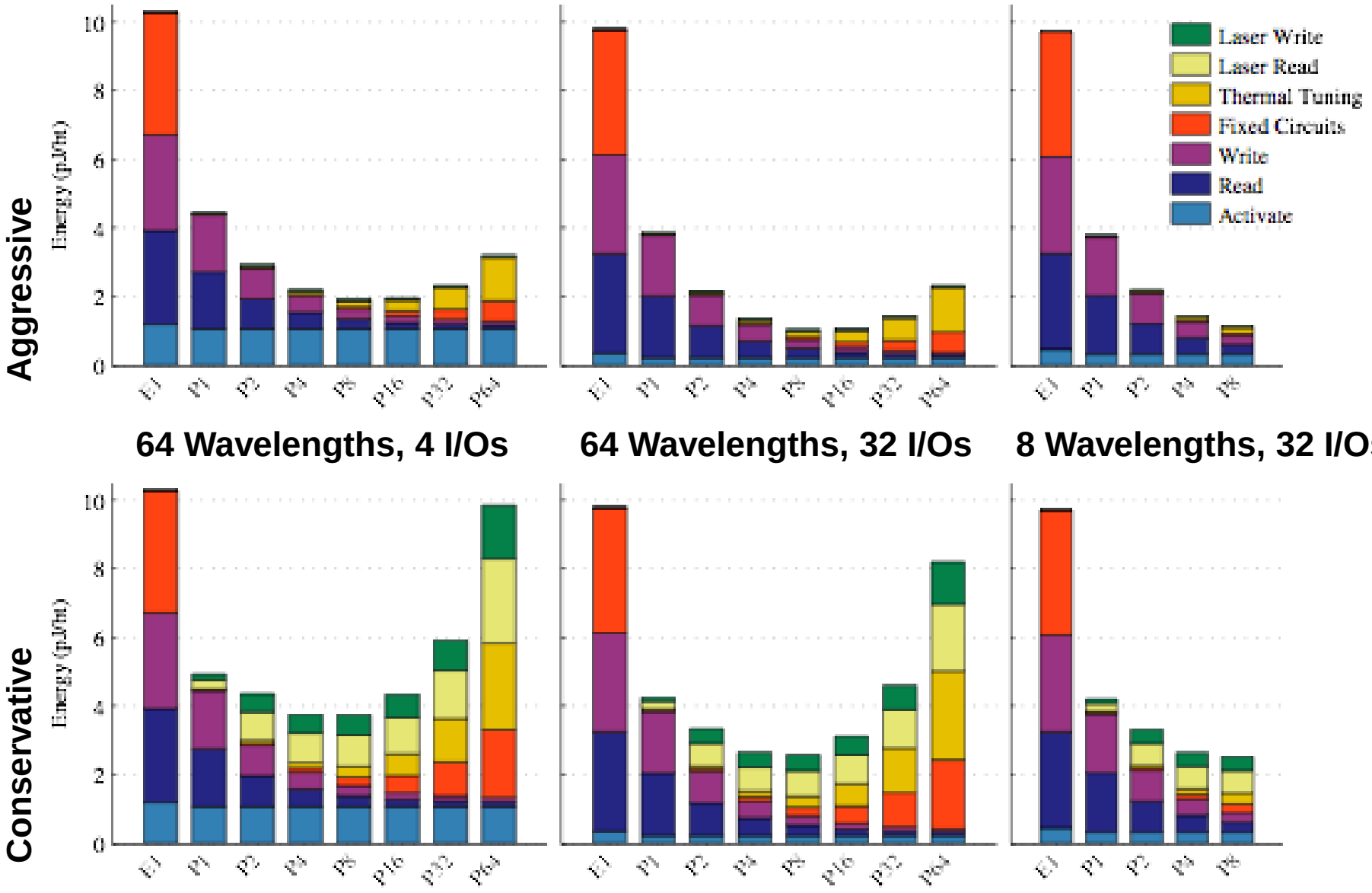
Electrical Baseline & Control H-Tree



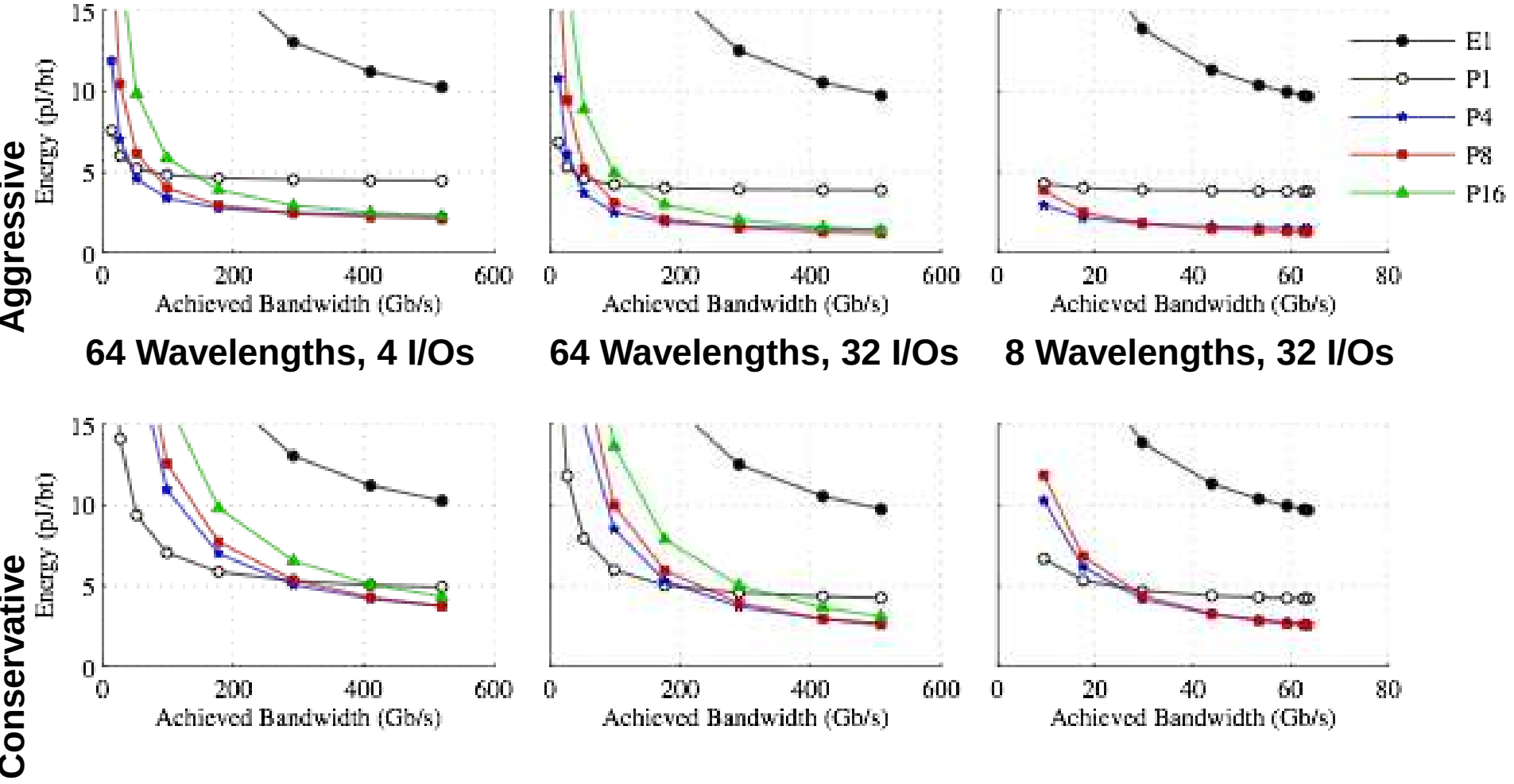
Photonic Floorplan showing Control Access Point

- Control distributed from the center of the chip
  - H-tree spreads out to banks
  - Can power gate control lines to inactive banks

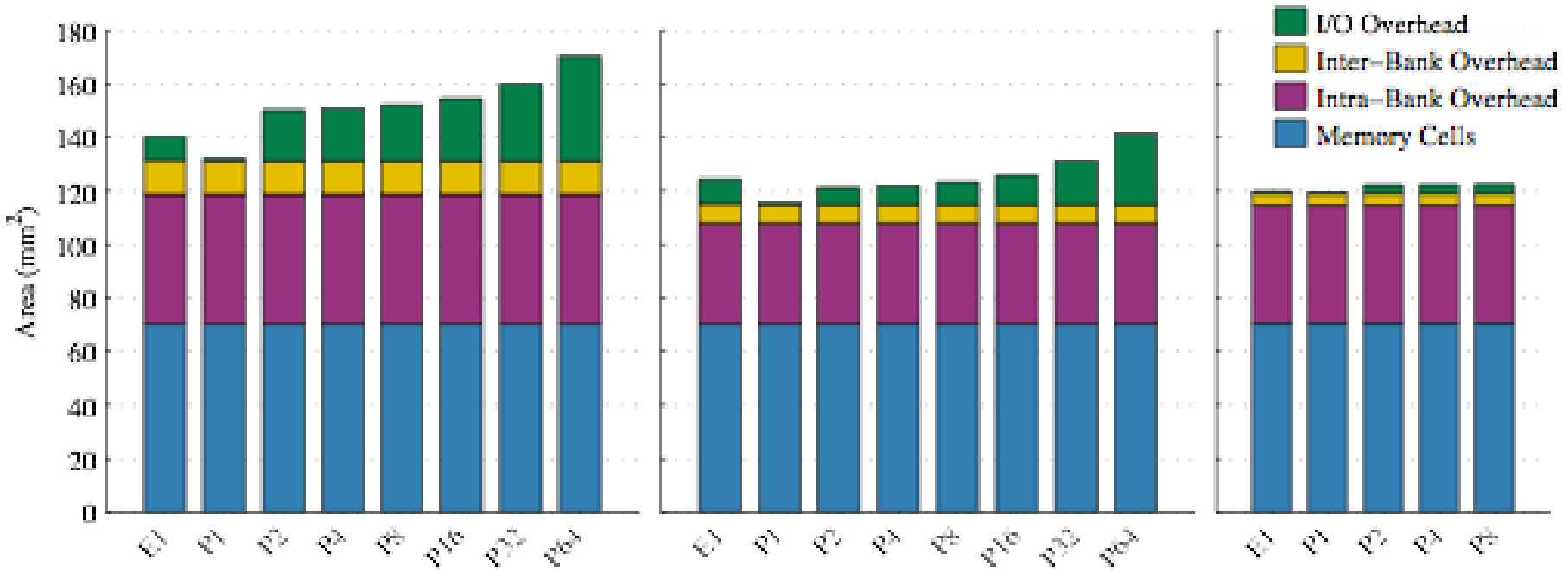
Full Energy


Aggressive
Conservative
Energy (pJ/bit)
0
2
4
6
8
10
Laser Write
Laser Read
Thermal Tuning
Fixed Circuits
Write
Read
Activate
64 Wavelengths, 4 I/Os
64 Wavelengths, 32 I/Os
8 Wavelengths, 32 I/Os

Utilization

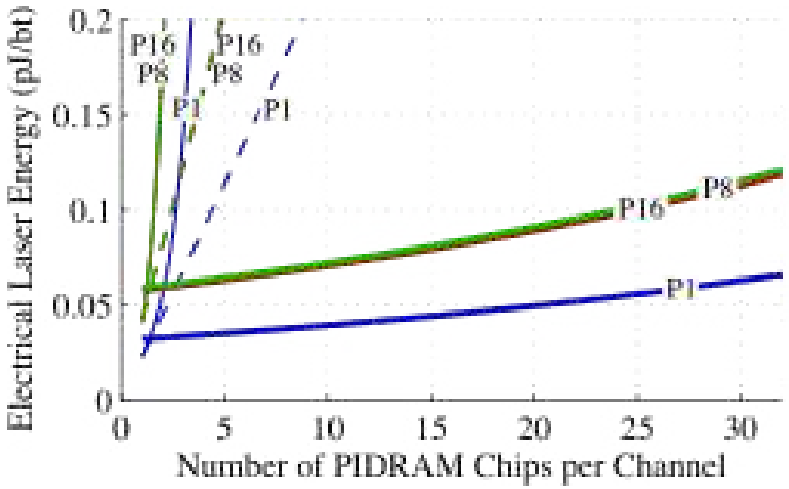


Full Area

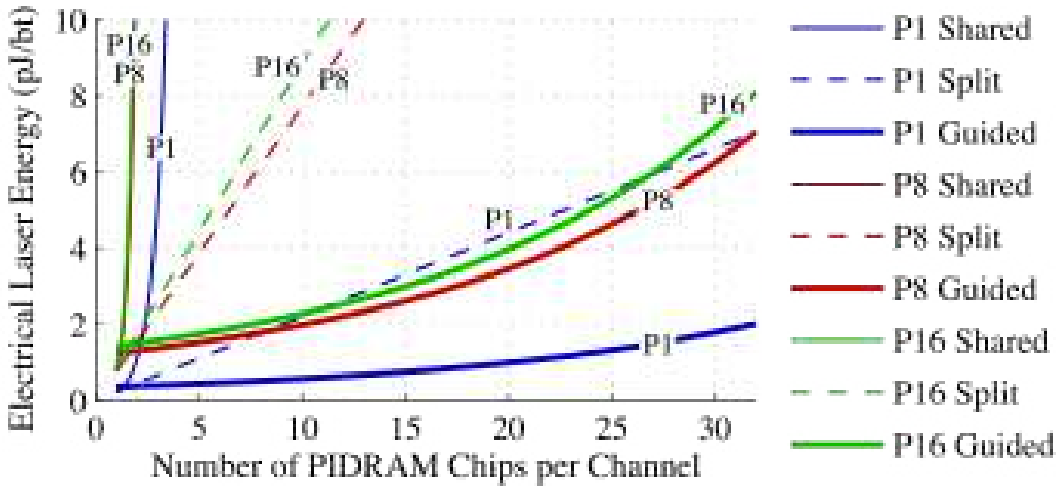


64 Wavelengths, 4 I/Os **64 Wavelengths, 32 I/Os** **8 Wavelengths, 32 I/Os**

Full Scaling



Aggressive



Conservative